Text-to-Speech in High-Variability Phonetic Training: Focus on L2 Phonological Awareness

Forcan Al-Shami[1*], Walcir Cardoso[2]

[1] Concordia University, Canada

[2] Concordia University, Canada

[*]Corresponding author's email: f_lshami@live.concordia.ca

[*]ORCID: https://orcid.org/0009-0003-6893-0220

Received: 28/02/2025 Revision: 23/04/2025 Accepted: 12/08/2025 Online: 25/12/2025

ABSTRACT

Time and space constraints in foreign/second language (L2) instruction often restrict learners' exposure to phonetic variability, a key factor in pronunciation development. High-Variability Phonetic Training (HVPT) offers a promising solution by exposing learners to phonetic variation; however, its implementation into instructional settings remains underexplored. This study investigates the integration of Text-To-Speech (TTS) technology with HVPT to provide varied L2 input in a semi-autonomous (beyond-the-classroom) environment. A mixed-methods pretest-posttest design examined discrete aspects of English pronunciation development, focusing on learners' phonological awareness of past *-ed* allomorphy. Thirty Arabic-speaking adult ESL learners in Kuwait were divided into a Treatment Group (exposed to varied TTS voices) and a Control Group (exposed to a single TTS voice), engaging in self-paced listening, categorization, and form-focused activities over four weeks. Results revealed significant improvements in phonological awareness for both groups, with no statistically significant difference between them. These findings contribute to ongoing debates about HVPT's added value in semi-autonomous settings and suggest that TTS technology alone—whether implemented with HVPT or not—can effectively support phonological awareness, offering a flexible and accessible tool for L2 pronunciation practice.

Keywords: High-Variability Phonetic Training (HVPT); Text-To-Speech (TTS); Second Language Acquisition; L2 Pronunciation; Phonological Awareness.

Introduction

Pronunciation is a fundamental yet often overlooked aspect of second language acquisition (SLA). While communicative competence encompasses grammar, vocabulary, and pragmatics, intelligible pronunciation is crucial for effective communication (Fraser, 2000; Prashant, 2018). However, traditional L2 instruction prioritizes grammar and vocabulary over pronunciation, limiting learners' exposure to diverse phonetic input (see Farhat & Dzakiria, 2017 and Jing, 2010 for similar claims). This issue is further exacerbated by classroom constraints, where learners primarily rely on the teacher's input due to limited instructional time and access to varied linguistic input (Bione & Cardoso, 2020; Collins & Muñoz, 2016).

Research in L2 pronunciation highlights that exposure to multiple speakers enhances learners' ability to generalize phonetic patterns across different voices, leading to more robust pronunciation skills (Levis, 2016; Logan et al., 1991). High-Variability Phonetic Training (HVPT) has improved the perception and production of challenging L2 sounds by exposing learners to diverse speech stimuli that vary in accents, voice types, and pitch (Levis, 2016). This phonetic variability fosters speech perception and facilitates the transfer of learning to novel linguistic contexts (Bradlow & Bent, 2008; Lively et al., 1993). However, HVPT has predominantly relied on recordings of multiple human speakers or synthesized voices, with most research conducted in controlled laboratory settings (Barriuso & Hayes-Harb, 2018). This raises concerns about its practical applicability in real-world classroom environments, where factors such as time constraints and resource limitations often restrict pronunciation training (Thomson, 2018). As a result, learners frequently receive inadequate exposure to phonetic variability (Collins & Muñoz, 2016; Thomson, 2018). Consequently, pronunciation training, particularly approaches incorporating HVPT, remains underutilized in many L2 learning contexts.

To address these limitations, recent advances in speech synthesis technology offer an innovative alternative for delivering varied phonetic input. Text-to-speech (TTS) technology, which converts written text into spoken output, enables learners to access diverse linguistic input essential for HVPT (Bione & Cardoso, 2020). While studies suggest that TTS can enhance pronunciation training and foster learner autonomy (Cardoso, 2018, 2022; Kiliçkaya, 2008), its integration within HVPT remains underexplored.

This study examines the role of TTS-assisted HVPT in improving ESL learners' phonological awareness of regular past tense allomorphy in English (i.e., past *-ed*; found in talk/t/, live/d/ and need/ɪd/). Specifically, the study compares the effectiveness of highly-variable TTS input (multiple synthetic voices) versus low-variability TTS input (a single synthetic voice) to determine whether exposure to greater phonetic variability yields superior pronunciation outcomes. By investigating TTS as a scalable tool for pronunciation training, this study contributes to the growing field of computer-assisted pronunciation training and offers insights for developing accessible, technology-enhanced pronunciation interventions. Findings from this research will also provide practical implications for educators seeking innovative, flexible alternatives to traditional pronunciation instruction that extend learning beyond the classroom.

Literature review

The Role of Input in Second Language Acquisition

The process of SLA is fundamentally shaped by the nature and quality of linguistic input learners receive. Rather than mere passive exposure, input involves comprehensible and meaningful interactions that facilitate learning, a concept central to Krashen's (1985) Input Hypothesis. According to this hypothesis, effective language acquisition occurs when learners are exposed to input slightly beyond their current proficiency level (i+1). However, exposure alone is insufficient; input must be comprehensible to promote learning. This has led educators and researchers emphasize the importance of linguistically rich environments that provide not only ample input but also high-quality, interactive learning experiences that enhance acquisition.

Research underscores the importance of both input quantity and quality in SLA. Flege (1995, 1999) found that learners with greater exposure to native speakers develop more accurate pronunciation and exhibit a reduced foreign accent. Similarly, Moyer (2009) highlights those high-quality interactions—particularly those involving meaningful communication with native speakers—lead to greater fluency and pronunciation accuracy. These findings suggest that while frequency of exposure is important, the authenticity and richness of input are equally crucial in shaping learning outcomes (Barcroft & Sommers, 2005).

A key factor in rich linguistic input is variability, which plays a crucial role in pronunciation learning (Bione et al., 2016). Studies have demonstrated that exposure to diverse phonetic models, incorporating differences in accent, speech rate, and intonation, enhances learners' ability to perceive and produce non-native phonemes. For example, research by Iverson and Evans (2009) demonstrated that learners trained with multiple accents and speech patterns developed stronger phonetic perception and production skills (see Thomson, 2012 for similar findings). These studies highlight the critical role of phonetic variability in developing robust pronunciation skills and facilitating phonetic generalization across linguistic contexts.

Despite the well-documented benefits of varied input, traditional classroom-based pronunciation training remains limited by instructional constraints, such as restricted time allocation and reliance on a single teacher's accent (Bione et al., 2016). These limitations hinder learners' ability to recognize and produce diverse phonetic contrasts, reinforcing the need for alternative instructional approaches that integrate structured and diverse linguistic models.

High-Variability Phonetic Training (HVPT) as an approach to Enriched Input

Given the crucial role of input in SLA, High-Variability Phonetic Training (HVPT) has been recognized as an effective approach for refining learners' phonetic perception and production skills. Unlike traditional pronunciation instruction, which often relies on single-speaker input, HVPT exposes learners to phonetic contrasts through diverse acoustic properties, such as multiple talkers, accents, speech patterns, and pitch variations (Ingvalson et al., 2014; Levis, 2016). This method integrates both inter-speaker variability, which captures differences in pronunciation styles across various speakers, and intra-speaker variability, which refers to fluctuations in speech within a single speaker due to contextual or emotional factors (Honeybone, 2011). By broadening learners' exposure to diverse phonetic stimuli, HVPT promotes phonetic generalization, allowing learners to apply learned phonetic distinctions across different communicative settings (Bradlow & Bent, 2008).

To incorporate phonetic variability into instruction, HVPT is typically implemented through perceptual training models, where learners listen to recorded speech samples from multiple speakers and engage in identification tasks with immediate feedback. These activities focus learners' attention on key acoustic properties of phonemes, helping them develop more precise phonetic categories (McCandliss et al., 2002). Immediate feedback plays a critical role in this process, as it enables learners to recognize distinctive acoustic cues, such as vowel formants, which are essential for intelligible pronunciation (Thomson & Derwing, 2016). This approach aligns with Schmidt's (1990) Noticing Hypothesis, which asserts that explicit awareness of phonetic contrasts is necessary for learners to integrate new linguistic input into their developing interlanguage system.

A substantial body of research has demonstrated the effectiveness of HVPT in L2 pronunciation training. Lively et al. (1993) found that Japanese learners who were trained on the English /l/-/r/ distinction through exposure to multiple speakers performed significantly better in perception tasks compared to those trained with a single talker. Moreover, the HVPT-trained group was able to generalize their learning to novel voices, suggesting that this method fosters phonetic adaptability. Similar outcomes have been observed in studies on Mandarin tone perception among English-speaking learners (Perrachione et al., 2011) and English vowel discrimination by Cantonese speakers (Wong, 2014), reinforcing HVPT's role in enhancing learners' ability to process unfamiliar phonetic input.

Further supporting evidence by Bradlow et al. (1997) highlights the broader implications of HVPT, showing that Japanese learners who underwent training to distinguish the English /r/-/l/ contrast not only improved in perception tasks but also demonstrated enhanced production accuracy. Additionally, a longitudinal study by Bradlow et al. (1999) revealed that HVPT-induced improvements persisted for at least three months post-training, indicating that phonetic gains acquired through HVPT are retained over time. A meta-analysis by Thomson (2018), which reviewed 32 studies on HVPT, further confirmed these findings, concluding that HVPT significantly enhances phonetic discrimination and production skills across diverse learner populations.

Despite its empirical support, HVPT remains underexplored in traditional L2 instruction due to several practical constraints. One major barrier is the difficulty of acquiring diverse speech samples, as HVPT depends on a wide range of talkers to ensure sufficient phonetic variability (Thomson, 2018). Additionally, integrating perceptual training within traditional classroom settings can be challenging, as it often requires specialized training software and a significant time investment from teachers. Given these challenges, emerging technologies, particularly through Computer-Assisted Language Learning (CALL), offer more dynamic and flexible ways to engage with the target language beyond formal classroom settings (Reinders & White, 2010).

A TTS-Assisted HVPT Approach to L2 Pronunciation Training

One such alternative for delivering HVPT outside the classroom is Text-To-Speech (TTS) technology, which has gained recognition as a valuable tool for supporting L2 pronunciation development (Cardoso, 2020, 2022). Unlike traditional classroom instruction, where pronunciation training is often constrained by time and limited exposure to diverse speech models, TTS provides learners with unlimited access to varied pronunciation input. This accessibility has positioned TTS as a significant asset in language learning, particularly for enhancing phonetic development (Cardoso, 2018; 2022; Kiliçkaya, 2008; Soler-Urzúa, 2011). Studies investigating its impact suggest that one of its strongest benefits is its ability to facilitate repeated exposure to different speech patterns, enabling learners to refine their pronunciation beyond the classroom (Ekşi & Yeşilçınar, 2016; González, 2007; Kim, 2018; Moon, 2012).

The effectiveness of TTS in language learning is supported by empirical research demonstrating its potential to enhance pronunciation skills (Bione et al., 2016; Bione & Cardoso, 2020; Liakin et al., 2017b; Soler-Urzúa, 2011). For instance, Liakin et al. (2017b) conducted a study in which learners using a TTS-based tool outperformed a control group in their ability to produce accurate French liaison. These findings reinforce the idea that controlled exposure to diverse phonetic input fosters better pronunciation accuracy. Other studies similarly emphasize that TTS supports phonetic learning by providing extended opportunities for self-directed practice, a key advantage in L2 acquisition (Bione et al., 2016; Soler-Urzúa, 2011).

A growing body of research also points to the potential benefits of integrating HVPT into computer-assisted pronunciation training. Thomson (2018) underscores the importance of varied phonetic input in L2 pronunciation training, a principle that aligns closely with the capabilities of TTS. Similarly, Bione and Cardoso (2020) highlight how TTS can simulate an HVPT environment by offering voice adjustments in terms of pitch, accent, and speaker identity. Such flexibility not only enhances learners' exposure to different speech models but also provides structured and individualized training that may not always be possible in traditional classroom settings.

This study builds on such insights by exploring how TTS can deliver the phonetic variability necessary for HVPT-based training, ultimately aiming to improve L2 pronunciation through a more accessible and adaptable learning approach.

Assessing English Past Tense -ed Allomorphy: Focus on Phonological Awareness

One of the primary phonological challenges for L2 learners is acquiring the past *-ed* allomorphy in English. The past *-ed* suffix is pronounced as /t/, /d/, or /ɪd/ depending on the preceding phonological environment. Specifically, /t/ follows voiceless consonants (e.g., talked [tɔkt]), /d/ follows voiced sounds (e.g., played [pleɪd]), and /ɪd/ occurs after alveolar stops /t/ and /d/ (e.g., needed [niːdɪd], wanted [wɒntɪd]).

For many ESL learners, particularly those whose first languages lack similar morphophonemic alternations, distinguishing and producing these allomorphic variations presents a significant pronunciation challenge (Cardoso, 2018; Collins et al., 2009). Arabic speakers, for example, frequently insert an epenthetic vowel in clusters deemed illicit in their L1, leading to mispronunciations such as asked being articulated as [æskɪd] (Kharma & Hajjaj, 1997). Additionally, the lack of transparent orthographic cues in English makes it difficult for learners to connect written forms with their correct phonetic realizations (Jackson & Cardoso, 2022).

To examine the acquisition of past *-ed* allomorphy in English, the study will focus on phonological awareness, the first stage of pronunciation development, as outlined in Celce-Murcia et al. (2010). According to the authors, pronunciation development follows a hierarchical progression, beginning with phonological awareness, advancing through aural perception and controlled oral production, ultimately leading to communicative oral proficiency. Phonological awareness involves learners' ability to identify and distinguish phonological structures such as segments and phonemic contrasts in the target language, all of which contribute to intelligible pronunciation (Anthony & Francis, 2005). According to Schmidt's (1990) Noticing Hypothesis, conscious awareness of phonetic distinctions is a prerequisite for acquisition, as learners must first perceive differences before accurately producing them. Similarly, Schmidt and Frota (1986) highlight the importance of recognizing mismatches between input and interlanguage to facilitate phonetic refinement. Research supports that explicit pronunciation training significantly enhances phonological awareness, reinforcing learners' ability to internalize and apply phonetic contrasts (Linebaugh & Roche, 2015).

Given its critical role in early-stage pronunciation development, this study prioritizes phonological awareness as the central focus for evaluating the effectiveness of TTS-assisted HVPT. TTS technology inherently emphasizes auditory input, making it a suitable tool for exposing learners to diverse phonetic contrasts (De Araújo Gomes et al., 2018).

Research Questions

Given the insufficient emphasis on pronunciation teaching in the ESL curricula (Barcomb & Cardoso, 2020), TTS technology presents an interesting pedagogical approach by delivering structured, varied phonetic input in an autonomous learning setting (Bione & Cardoso, 2020). By incorporating HVPT principles, TTS-assisted training has the potential to enhance phonological awareness and sensitivity to the morphophonemic variation of past *-ed*. Thus, this study addresses the following research question:

- How does TTS-assisted HVPT affect ESL learners' phonological awareness of English past tense *-ed* allomorphy?

Methods

Participants

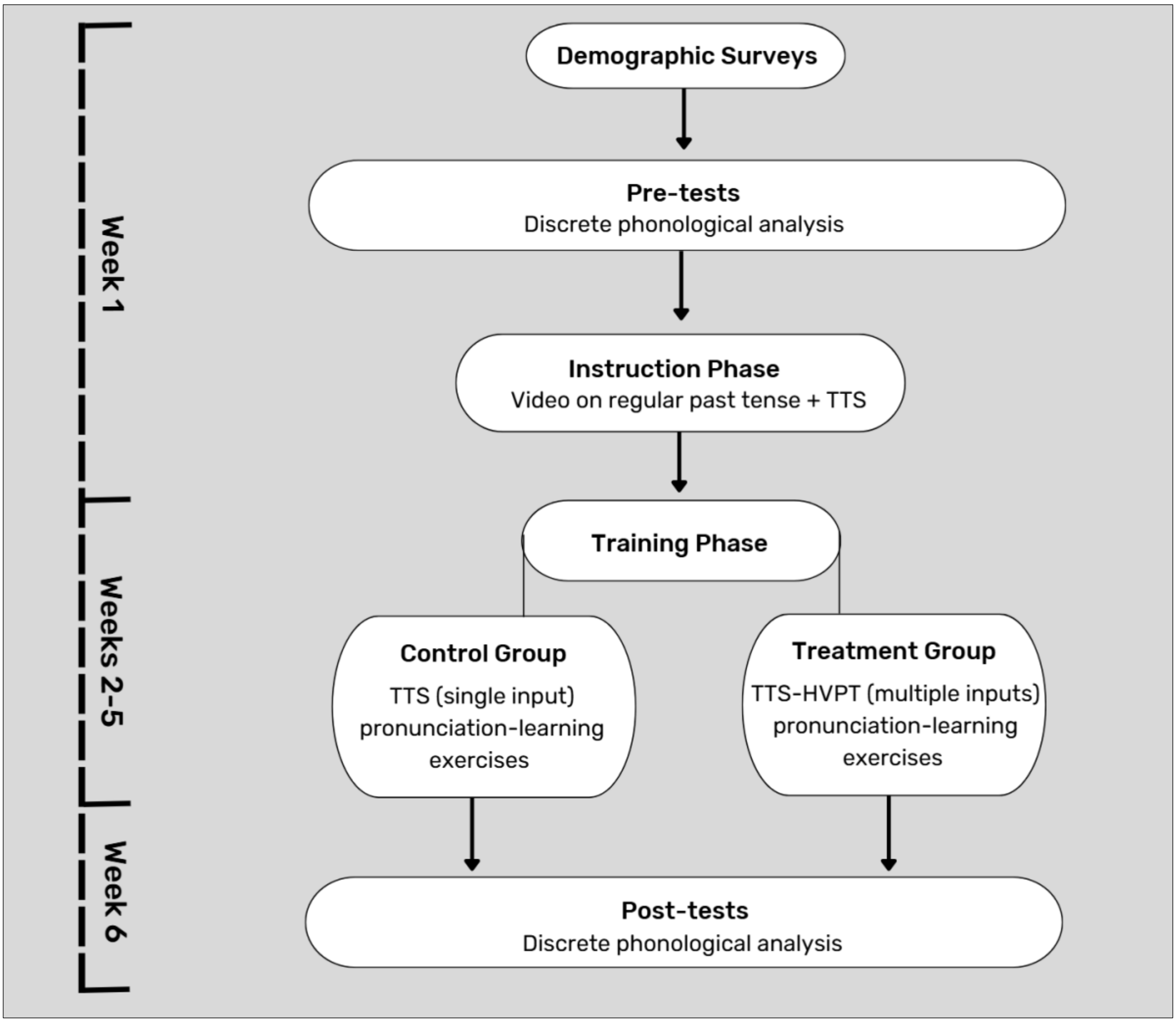
This study involved 30 adult ESL learners (9 males and 21 females) residing in Kuwait, aged between 18 and 30. All participants were native Arabic speakers with low to intermediate English proficiency and no reported history of speech or hearing disorders. Their proficiency was assessed through a combination of pre-test results (only those scoring 50% or below were included), self-assessment surveys, and pre-study interviews. Participants were randomly assigned to one of two groups: a treatment group (TTS-with HVPT) and a control group (TTS-with a single voice), each consisting of 15 learners. To preserve anonymity, each participant was assigned a unique identification code (e.g., P01, P02). Recruitment was conducted through social media announcements and outreach to English teachers within local academic institutions. The study was approved by Concordia University's Human Research Ethics Committee, and informed consent was obtained from all participants, who also received a small incentive equivalent to CAD $30 for their participation.

Research Design

This study followed a mixed-methods pretest-posttest design, conducted remotely over six weeks using Zoom and the FlexiQuiz platform. As outlined in Figure 1, the study was structured into distinct phases.

Figure 1.

Design of the study



In the initial phase, participants completed demographic surveys to gather essential information such as age, linguistic background, and familiarity with technology. Following this, they watched an instructional video introducing the English regular past tense and the written form of its inflection (i.e., the *-ed* suffix at the end of the verbs). To draw learners' attention to the sound–spelling relationship targeted in the study, pronunciation rules were intentionally withheld at this stage. In line with data-driven learning (e.g., Johns, 1991; Pérez-Paredes & Boulton, 2025), this guided-induction step encouraged learners to observe the input, form hypotheses about *-ed* realizations, and test these against additional examples. This strategy promoted noticing, a key precursor to phonological awareness development (Schmidt, 1990; Celce-Murcia et al., 2010). The introductory video also provided step-by-step guidance on using Speechify, the selected TTS tool, for pronunciation practice.

To minimize technical difficulties, participants had access to technical support via email and WhatsApp with the researcher. All participants completed a test run on the platform before beginning formal training to ensure familiarity with the digital tools.

Data collection and Assessment

To assess phonological awareness of past *-ed* allomorphy, participants completed pre-tests and post-tests, each lasting approximately 30 minutes. Two assessment instruments were used to evaluate learners' ability to recognize and categorize past *-ed* pronunciation. The first instrument consisted of open-ended phonological awareness questions, in which participants responded to conceptual questions assessing their understanding of past *-ed* pronunciation (e.g., *"Do you think the pronunciation of past tense -ed differs across different verbs?").* The second instrument was a phonological categorization task, in which participants listened to a set of -ed inflected verbs and classified them according to their perceived phonetic realization as /t/, /d/, or /ɪd/.

To ensure reliability in data collection, identical pre-test and post-test items were used but presented in a randomized order to minimize familiarity effects. This controlled for potential biases, ensuring that observed improvements were due to training rather than test repetition.

Training Phase

Participants engaged in ten self-paced training sessions, each lasting 30 minutes, over a period of four weeks. The control group was exposed to a single standardized TTS-generated native English voice throughout all sessions. The treatment group was exposed to multiple English accents via five distinct synthesized voices (3 females, 2 males), representing North American, British, and Australian English. While exact voice specifications (e.g., rate or accent subtype) were not available for each voice on the TTS platform, they reflected the target regional varieties. In addition, as the authors are not native speakers of these varieties, it was not possible to verify the detailed accent attributes of each voice.

The inclusion of five different voices aligns with HVPT research, which suggests that exposure to multiple talkers enhances phonetic generalization (Zhang et al., 2021; Lively et al., 1993; 1994; Sadakata & McQueen, 2014). Table 1 outlines the distribution of TTS-generated voices across training sessions.

Table 1.

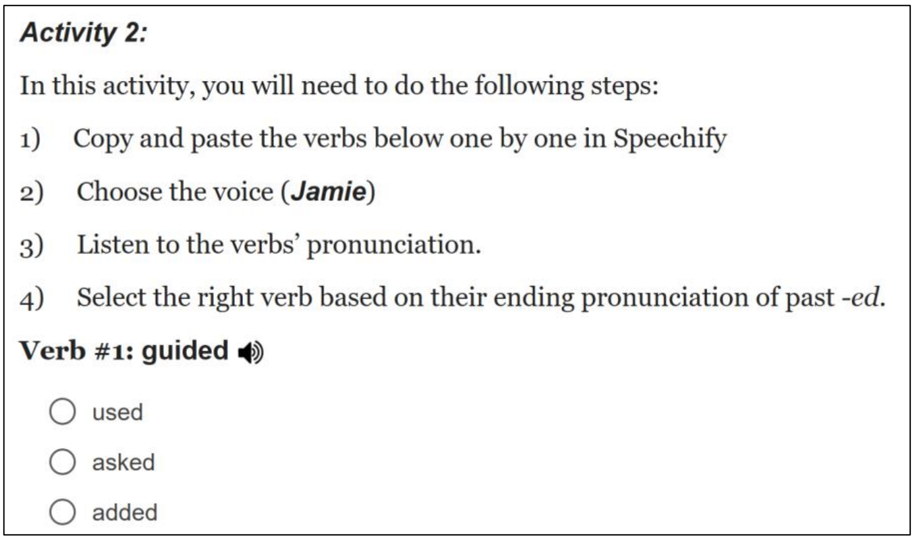
Session Distribution of TTS Voices

Treatment Group (TTS-with HVPT)			Control Group (TTS-single voice)		
Voice Type	Voice Attributes	Assigned Sessions	Voice Type	Voice Attributes	Assigned Sessions
Jamie	North American, Female	1 and 6	Jamie	North American, Female	All sessions
Micheal	British, Male	2 and 7			
Sydney	Australian, Female	3 and 8			
Nate	North American, Male	4 and 9			
Stephanie	British, Male	5 and 10			

Each training session incorporated three key TTS-assisted activities, including listening tasks, fill-in-the-blank exercises, and multiple-choice questions. Participants were instructed to copy and paste each given text into Speechify, select the designated TTS voice, and listen to the audio without reading the original text. Afterward, they answered comprehension questions and engaged in form-focused tasks, such as categorizing past *-ed* allomorphs by pronunciation, as illustrated in Figure 2. To enhance learning outcomes, immediate feedback was integrated into all tasks, allowing participants to receive real-time corrections, thereby reinforcing the principles of HVPT-driven phonological learning.

Figure 2.

Training Activity 2: A Sample



Activity 2:

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

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

Verb #1: guided

- used
- asked
- added

Note. A sample categorization task for the /ɪd/ allomorph, in which learners identified past *-ed* allomorphs by categorizing verbs based on their ending pronunciations.

This study leveraged TTS technology to introduce both inter-speaker and intra-speaker variation in pronunciation training. The primary focus was on inter-speaker variability, exposing learners to multiple TTS-generated voices with different speech patterns and accents, including North American, British, and Australian English. While intra-speaker variation, such as speed modifications within a single voice, could also contribute to phonetic flexibility, this study prioritized exposure to multiple speakers to enhance learners' adaptation to diverse pronunciation models. This approach aligns with the English as a Lingua Franca perspective (Jenkins, 2000), which advocates for preparing learners to understand and produce diverse English varieties.

Data Analysis

Given that our outcome measures included a 4-point ordinal scale for phonological awareness and bounded categorization scores (0–21; seven items per allomorph), and considering the relatively small sample size in each group (n = 15), we employed non-parametric methods for analysis. Such tests are well-suited for data where equal intervals between values cannot be assumed and where distributions may be non-normal or skewed, which is often the case with discrete, ceiling-limited scores. Accordingly, within-group changes from pre- to post-test were analyzed using the Wilcoxon Signed-Rank Test to determine whether individual learners showed statistically significant improvements in phonological awareness after training. Between-group comparisons were conducted using the Mann–Whitney U Test to assess whether differences between the treatment and control groups were statistically significant. Finally, to control for Type I errors resulting from multiple comparisons, Holm–Bonferroni correction was applied to adjust p-values, ensuring the validity of statistical inferences.

Phonological awareness was measured using two assessment instruments. The first assessment (open-ended questions task) was evaluated using a four-point rating scale to assess learners' knowledge of past *-ed* pronunciations. A score of 1 indicated no awareness regarding past *-ed* pronunciation, while a score of 2 reflected partial recognition, with learners correctly identifying only one allomorph. A score of 3 denoted recognition of two of the three allomorphic forms, and a score of 4 was assigned to participants who successfully identified all three allomorphs. The second assessment (phonological categorization task) measured learners' ability to classify past *-ed* allomorphs using a 21-point scale (seven items to each allomorphic category). These measures provided a comprehensive assessment of learners' phonological awareness, capturing their ability to distinguish and generalize past *-ed* pronunciation patterns following the intervention.

Results

This study examined the impact of TTS-assisted HVPT on ESL learners' pronunciation, focusing on discrete phonological analysis, particularly phonological awareness of past-tense -*ed* allomorphy. The analysis compared two training approaches: exposure to multiple English accents versus exposure to a single native accent, via TTS technology. The results include findings from the two assessment instruments, with statistical analyses and interpretations detailed below.

Phonological Awareness 1: Open-Ended Responses

The first measure of phonological awareness assessed participants' conceptual knowledge of English past *-ed* pronunciation through open-ended questions (e.g., *How do you think the past -ed is pronounced? Explain your answer.)* Participants were rated on a four-point scale: 1 = no awareness, 2 = recognition of one allomorph, 3 = recognition of two allomorphs, and 4 = full awareness of all three allomorphs.

Within-Group Analysis

Results from the Wilcoxon Signed-Rank Test indicated a statistically significant improvement in the treatment group, with the median increasing from 3 to 4 following training. The control group also showed a significant pre–post improvement ($p = 0.0005$; Holm-Bonferroni-corrected $p = 0.0001$), although the change was less pronounced. In this case, the median remained at 3 before and after training, and the significance instead reflected a shift in the distribution of scores, more and/or larger positive differences than negative ones, rather than a change in central tendency. This illustrates how the Wilcoxon test can detect subtle distributional changes that median-based summaries may overlook, highlighting the difference between statistical significance and central tendency measures.

Between-Group Analysis

Results from the Mann-Whitney U Test revealed no statistically significant difference in improvement between the treatment and control groups. ($z = 144.0$, $p = 0.1104$, corrected $p = 0.1104$). These findings suggest that while the treatment group demonstrated greater gains in phonological awareness, reflected in a median increase from 3 to 4 (see Table 2), the difference between groups did not reach statistical significance. A summary of these results is provided in Table 2, where statistically significant differences are marked in bold.

Table 2.

Results Summary for Phonological Awareness 1

Measure	Group Type	z-statistic test	p-value		Median	
			Uncorrected	Corrected	Pre test	Post test
Awareness 1 (open-ended responses)	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: Categorizing Inflected -ed Forms

The second phonological awareness assessment measured participants' ability to match 21 target *-ed* forms (seven per allomorph) with their respective allomorphic pronunciations.

Within-Group Analysis

The Wilcoxon Signed-Rank Test indicated statistically significant improvements across all allomorphs in the treatment group. For the /t/ allomorph, scores increased from 2.27 pre-training to 5.07 post-training ($z = 0.0$, $p > 0.0001$, corrected $p = 0.0005$). Similarly, the /d/ allomorph showed notable progress, with median scores rising from 1.67 to 5.80 ($z = 0.0$, $p > 0.0001$, corrected $p = 0.0005$). The /ɪd/ allomorph also exhibited significant improvement, with scores increasing from 5.20 to 6.93 ($z = 0.0$, $p \geqslant 0.0008$, corrected $p = 0.0049$).

While the control group also demonstrated statistically significant gains, the effect sizes were smaller. For the /t/ allomorph, scores increased from 2.4 to 4.4 ($z = 2.5$, $p = 0.0023$, corrected $p = 0.0092$). The /d/ allomorph saw an increase from 1.6 to 4.6 ($z = 0.0$, $p > 0.0001$, corrected $p = 0.0005$), while for /ɪd/, scores improved from 4.27 to 6.8 ($z = 0.0$, $p = 0.0009$, corrected $p = 0.0049$). Although the control group made notable progress, the treatment group consistently demonstrated larger improvements across all allomorphic categories.

Between-Group Analysis

Comparisons using the Mann–Whitney U Test revealed no statistically significant differences between the treatment and control groups for the /t/ ($z = 141.0$, $p = 0.2306$, corrected $p = 0.2783$) and /ɪd/ ($z = 77.5$, $p = 0.1391$, corrected $p = 0.2783$) allomorphs. For /d/, the uncorrected p-value was below 0.05 ($z = 163.5$, $p = 0.0262$), indicating a trend toward improvement. However, after applying the Holm–Bonferroni correction (corrected $p = 0.0788$), this effect did not reach statistical significance. This suggests that, although the treatment group showed numerically greater improvement for /d/, the difference cannot be considered statistically reliable.

Overall, the findings from the second phonological awareness assessment indicate that the treatment group demonstrated improvements across all allomorphs, with the most notable gains observed for /t/ and /d/. While the control group also showed progress, the extent of improvement was relatively smaller. Although no statistically significant differences were detected between the groups for /t/ and /ɪd/, the results revealed a trend toward greater improvement in /d/ among treatment group participants. Table 3 presents the descriptive and inferential statistics for Phonological Awareness 2, while Figures 3, 4, and 5 provide graphical representations of performance across the /d/, /t/, and /ɪd/ allomorphs.

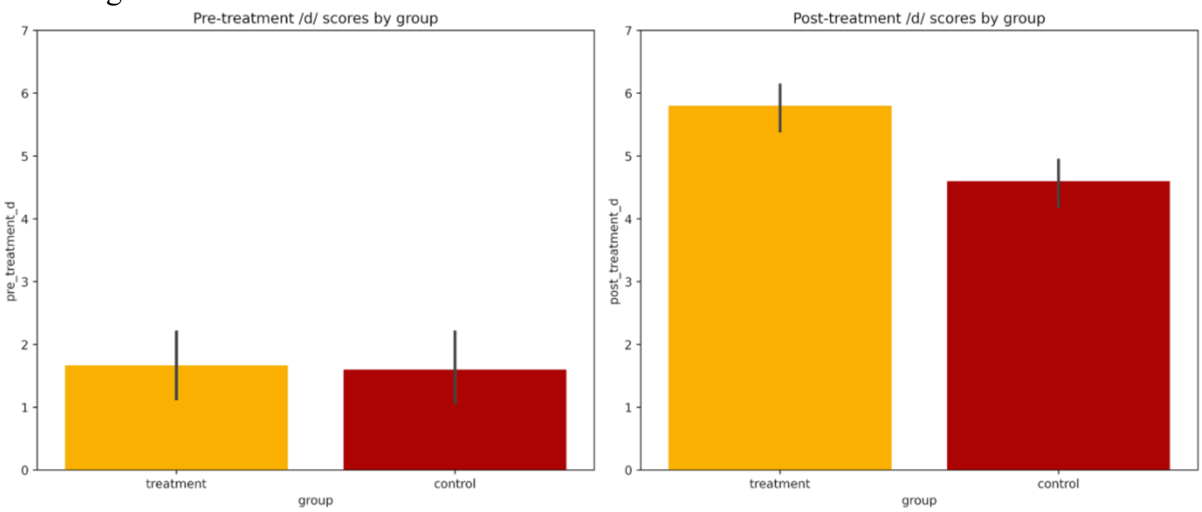
Table 3.

Results Summary for Phonological Awareness 2

Measure (Allomorph)	Group Type	z-statistic test	p-value		Mean	
			Uncorrected	Corrected	Pre test	Post test
/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	-	-
/ɪd/	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	-	-

Figure 3.

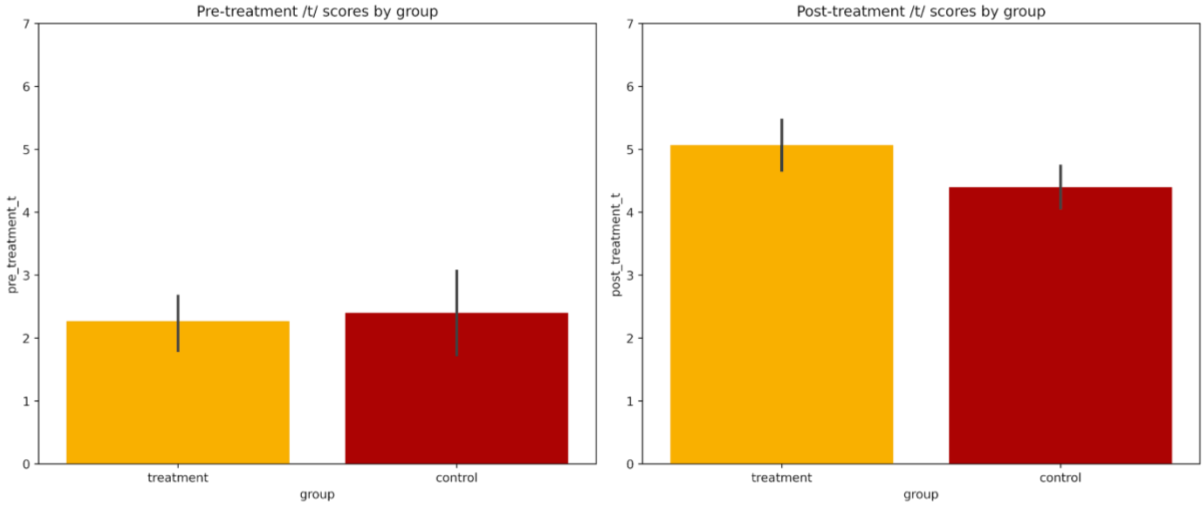
Phonological Awareness 2: /d/



Note. Mean pre- and post-test scores for /d/ phonological awareness by group.

Figure 4.

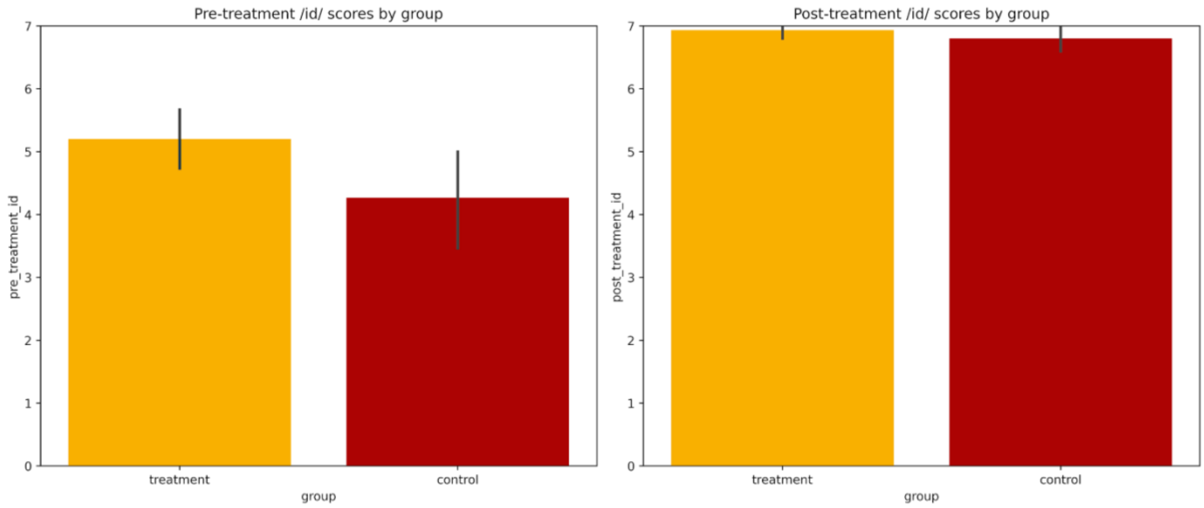
Phonological Awareness 2: /t/



Note. Mean pre- and post-test scores for /t/ phonological awareness by group.

Figure 5.

Phonological Awareness 2: /ɪd/



Note. Mean pre- and post-test scores for /ɪd/ phonological awareness by group.

Discussion

This study investigated the impact of TTS-assisted HVPT on ESL learners' phonological awareness of past *-ed* allomorphy. The primary objective was to assess whether high-variability phonetic training through TTS technology enhances learners' ability to identify the morphophonemic variations of English past *-ed*. The findings indicate that while participants exposed to HVPT with multiple TTS-generated voices demonstrated greater improvements, the differences between the treatment and control groups were not statistically significant. This suggests that TTS-based pronunciation training, regardless of its integration with HVPT, can effectively support phonological awareness development.

Phonological Awareness Gains and the Role of TTS

Results from both conditions showed significant improvements in participants' awareness of past *-ed* allomorphs, aligning with previous research demonstrating the effectiveness of TTS-based pronunciation training in fostering phonological awareness, even in the absence of high-variability input (Bione & Cardoso, 2020; Cardoso, 2018). While the treatment group exhibited greater gains, the lack of statistically significant differences between groups suggests that TTS-generated input, whether from a single voice or multiple voices, was sufficient for phonological learning.

The within-group improvements observed in the HVPT condition further support prior findings on the advantages of phonetic variability in pronunciation training (Thomson, 2018). However, the fact that the control group also showed progress, coupled with the absence of statistically significant between-group differences across the three allomorphs, suggests that TTS technology, whether incorporating a single voice or HVPT, offers effective phonological input that facilitates learners' pronunciation development. This outcome may be attributed to the intelligibility, consistency, and naturalness of synthesized speech, which previous studies have identified as a valuable tool for pronunciation instruction (Cardoso et al., 2015), particularly in addressing perceptually challenging phonetic contrasts (Bione et al., 2017; John & Cardoso, 2017). In this sense, the pedagogical affordances of TTS—namely, its clear and consistent input, combined with opportunities for autonomous access—may have compensated for the lack of variability in the control group, thereby narrowing the performance gap between variable and non-variable input conditions.

Comparing Allomorphic Awareness: Greater Gains in /d/

When examining individual allomorphs, /d/ exhibited the most substantial improvement in the treatment group, suggesting that high-variability input may be particularly beneficial for acquiring this form. The control group also showed progress, albeit to a lesser extent. One possible explanation is the strong orthographic connection between the -ed spelling and the /d/ pronunciation, which may have facilitated its acquisition (Delatorre, 2010).

Although no statistically significant differences emerged between groups, the trend toward greater improvement in the HVPT condition for /d/ is consistent with studies highlighting the advantages of high-variability input in phonological learning (Bradlow & Bent, 2008; Thomson, 2018). While HVPT appears to enhance phonological awareness, these findings also emphasize that a single, high-quality TTS-generated voice can be as effective in pronunciation training.

Challenges in Acquiring /t/

In the initial assessments, /t/ was the least recognized allomorph, with most participants in both groups failing to identify it in the first phonological awareness test. However, post-training results showed significant improvements in /t/ awareness across both groups, though the gains were less pronounced compared to /d/. These findings are consistent with previous research indicating that certain allomorphic variations are more resistant to acquisition due to their articulatory and/or perceptual complexity. For example, Cardoso (2018) found that English L2 learners struggled to acquire /t/ because of its sonority profile: being less sonorous than the other allomorphs /d/ and /ɪd/, /t/ is less aurally perceptible and consequently more resistant to acquisition.

Another possible reason for the lower gains in /t/ recognition is the lack of a direct orthographic correspondence between the spelling *-ed* and the /t/ pronunciation. Unlike /ɪd/ and /d/, which maintain a clear connection with the written form *-ed*, /t/ has no evident grapheme-to-sound relationship, making it more challenging for learners to identify (see also Cardoso, 2018; Dwight, 2012 for similar claims). Jackson and Cardoso (2022) argue that inconsistencies between L1 and L2 orthographic systems often create additional challenges in phonological acquisition. As a result, the opaque nature of the orthographic cues for /t/ likely contributed to the difficulties learners faced in acquiring this allomorph.

Minimal Gains in /ɪd/

In comparison to /t/ and /d/, the /ɪd/ allomorph exhibited the least improvement from pre-test to post-test scores. This was likely due to a ceiling effect, as learners already demonstrated high initial awareness of this allomorph prior to training. Because /ɪd/ maintains a transparent relationship with the -ed spelling, participants likely recognized it more readily, leaving little room for further improvement. This pattern aligns with previous findings, such as those of Delatorre (2010), who highlights the facilitative role of transparent orthography in the acquisition of past *-ed* forms.

Additional factors may have influenced the relative ease of acquiring the /ɪd/ allomorph. Research on L1 phonological transfer suggests that languages restricting complex coda clusters, such as Arabic, may impact how learners process past *-ed* in English (Barros, 2003; Zimmer et al., 2009). For Arabic-speaking English learners, certain consonant clusters, such as /gd/ in "arran/gd/", are often avoided, prompting the insertion of an epenthetic vowel to simplify articulation (Kharma & Hajjaj, 1997; Salim & Mohammed, 2023). This process may explain why Arabic speakers naturally gravitate toward an /ɪd/ realization, even in contexts where /d/ or /t/ would be expected.

Conclusion

This study examined the effectiveness of TTS-assisted HVPT in improving ESL learners' phonological awareness of past *-ed* allomorphy. The findings suggest that while exposure to high-variability phonetic input may offer certain pedagogical benefits, TTS technology alone, regardless of input variability, plays a significant role in supporting phonological awareness development. Overall, these findings align with previous research that highlights the quality and consistency of synthetic speech as key factors in facilitating phonological development (see De Araújo Gomes et al., 2018 for similar claims).

From a pedagogical perspective, these findings offer valuable insights for L2 pronunciation instruction. The observed improvements in phonological awareness across both training conditions highlight the potential of TTS technology as a flexible and accessible tool for pronunciation learning, particularly in contexts where access to native-speaker input is limited. By integrating TTS-assisted HVPT into instructional settings, educators can provide learners with varied auditory input that reinforces phonological distinctions and minimizes L1 interference. For example, teachers can incorporate TTS into their lessons through targeted homework drills, such as practicing /t/ and /d/ distinctions with fill-in-the-blank or minimal pair activities. In class, these activities can be complemented by communicative tasks, where the teacher provides personalized feedback, helping learners apply their phonological awareness in more meaningful, interactive contexts. The ability to customize and control phonetic input through TTS further enhances its utility for autonomous learning, allowing learners to engage with pronunciation training beyond the constraints of classroom instruction.

Beyond its pedagogical applications, this study contributes to the growing body of research on speech technology in L2 education, emphasizing the importance of incorporating innovative tools like TTS alongside evidence-based pronunciation training methods such as HVPT. The results support the integration of synthetic speech into pronunciation pedagogy, offering learners increased opportunities for auditory exposure and self-directed practice.

Despite its contributions, this study has several limitations that require consideration in future research. First, while phonological awareness was the primary focus of this research, pronunciation development is a multi-faceted process that also involves aural perception and oral production (Celce-Murcia et al., 2010). A more comprehensive evaluation, including both controlled and spontaneous speech production, would provide a fuller understanding of the impact of TTS-assisted HVPT on overall pronunciation skills. Second, the short duration of the study presents another limitation. The training period spanned only one month, which may have restricted the extent of phonological improvement observed in participants. Extending the training duration and incorporating spaced practice over a longer period would likely yield a more robust understanding of the long-term benefits of TTS-assisted HVPT. Future research should adopt a longitudinal approach to explore the sustainability of pronunciation gains following prolonged exposure to high-variability TTS input. Third, we could not independently verify the attributes or accent authenticity of the TTS voices beyond the platform's labels. Without technical specifications (e.g., speaking rate, pitch range) or acoustic validation by experts, the actual phonetic variability may have been lower or higher than intended. Future studies should verify and report voice characteristics, or use TTS systems with documented metadata. Lastly, the study was conducted with a relatively small sample size within a specific linguistic and cultural context: Kuwaiti ESL learners. While the findings provide valuable insights into pronunciation development in this learner group, their generalizability remains limited. Future studies should expand the participant pool, incorporating learners from diverse linguistic backgrounds and varied educational contexts to explore the broader applicability of TTS-assisted HVPT.

By addressing these limitations, future research can further refine the role of TTS-assisted HVPT in pronunciation training, exploring its potential to enhance not only phonological awareness but also speech perception and production across different L2 learning environments.

Acknowledgments

We sincerely appreciate the contributions of those who made this study possible. Our deepest gratitude goes to Dr. Paul John for his invaluable feedback, as well as to Damian Diaz for his expert assistance with the statistical analysis. We are especially grateful to the participants, whose dedication and willingness to engage in this research were essential to its success. Finally, we acknowledge financial support from the Social Sciences and Humanities Research Council (SSHRC; Cardoso: 430-2022-00512).

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

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

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.

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

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).

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

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.

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

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

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

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

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

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.

Pérez-Paredes, P., & Boulton, a. (2025). *Data-driven learning in and out of the language classroom*. Cambridge University Press.

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.

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

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.

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.

Biodata

Forcan Al-Shami is a PhD student in the Department of Education at Concordia University, Canada. Her research focuses on exploring the integration of text-to-speech (TTS) technology with high-variability phonetic training (HVPT) to enhance L2 pronunciation. Since 2011, she has specialized in teaching English as a second language to Arabic speakers in Kuwait.

Walcir Cardoso is a Professor of Applied Linguistics at Concordia University, Canada. He conducts research on the L2 acquisition of phonology, morphosyntax and vocabulary, and the effects of computer technology (e.g., clickers, text-to-speech synthesizers, automatic speech recognition, intelligent personal assistants) on L2 learning.