



AI-Generated Versus Human-Crafted Songs and EFL Learners' Pronunciation Accuracy

Arghavan Mansouri Rad ^{1,*}, Mahnaz Mostafaei Alaei ²

^{1,*}Corresponding author, Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran; arghavan1378arghavan@gmail.com

²Department of English Language and Literature, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran; ommmostafaei@atu.ac.ir

Original Research Article

Date of Submission: 10 May 2026

Date of Acceptance: 25 June 2026

Abstract

Pronunciation is a challenging aspect of English language learning, and songs have been used as pedagogical resources in pronunciation instruction. More recently, artificial intelligence (AI) has enabled the creation of customized songs for language-learning purposes. This study compared pronunciation accuracy among Iranian English as Foreign Language (EFL) learners who received instruction incorporating AI-generated songs and those who used human-crafted songs. The study employed a pretest–posttest quasi-experimental comparison-group design involving 42 upper-intermediate Iranian EFL learners (20 in the AI-generated-song group and 22 in the human-crafted-song group). Pronunciation accuracy was assessed before and after the instructional intervention. Both groups demonstrated significant pretest-to-posttest improvements; however, an analysis of covariance (ANCOVA), controlling for pretest scores, revealed a statistically significant between-group difference in posttest pronunciation accuracy, $F(1, 39) = 6.13$, $p = .01$, partial $\eta^2 = .15$, favoring the AI-generated-song group. These findings indicate that the use of AI-generated songs was associated with greater gains in pronunciation accuracy than the use of human-crafted songs in the context investigated. Nevertheless, the quasi-experimental design and relatively small, context-specific sample limit causal inference and generalizability. The findings offer preliminary evidence for considering customized AI-generated songs as supplementary resources in EFL pronunciation instruction, while further research is needed to examine the generalizability and durability of the observed gains.

Keywords: Adaptive Learning, AI-Generated Song, Artificial Intelligence, Human-Crafted Song, Pronunciation Accuracy

1 Introduction

Within the increasingly demanding landscape of English as a Foreign Language (EFL) education, scholars have prioritized the exploration of teachers' psychological and pedagogical attributes. Central to this inquiry are Emotional Intelligence (EI), Teaching for Creativity (TFC), and Work Engagement (WE), which are frequently studied for their influence on pedagogical quality and professional efficacy. Emotional Intelligence, conceptualized as the capacity to identify, interpret, and manage emotional states (Salovey & Mayer, 1990), is extensively documented as a key determinant of

professional performance. Similarly, work engagement is defined by distinct components of vigor, dedication, and absorption (Schaufeli et al., 2002), with prior studies, such as Yang (2022), indicating that EFL instructors' emotional resources often correlate with their level of engagement, though causality remains unproven.

Music, often referred to as a universal language, has consistently played a vital role in language learning. Among its various forms, songs have proven particularly effective in engaging learners and fostering language acquisition, often outperforming traditional instructional methods (Ludke, 2009). In the context of English as a Foreign Language (EFL), exposure to English songs is a powerful tool for developing linguistic skills, especially in speaking and pronunciation, which serve as the foundation of oral communication. As Levis (2018) states, "pronunciation is an important sub-system of language, without which successful communication is not achieved" (Alghazo et al., 2023, p. 1).

Besson et al. (2011) discuss important links between music and language processing; however, this relationship alone does not explain how musical activities may support L2 pronunciation development. The present study is additionally informed by the Revised Speech Learning Model (SLM-r), which proposes that L2 phonetic learning is influenced by the quantity and quality of input and by learners' perception of differences between L1 and L2 sounds (Flege & Bohn, 2021). From this perspective, songs may provide repeated exposure to English pronunciation patterns and opportunities to notice and reproduce target sounds. Their rhythmic and melodic structures may also draw learners' attention to word stress, timing, and intonation, although these potential benefits require empirical examination. Both AI-generated and human-composed songs may provide such opportunities; however, differences in their linguistic content and instructional suitability could influence learning outcomes. Accordingly, the present study compares the effects of AI-generated and human-composed songs on EFL learners' pronunciation accuracy, including segmental and suprasegmental features. The SLM-r thus provides a theoretical basis for considering the role of phonetic input in pronunciation learning, while the comparative design examines whether the two song-based instructional approaches are associated with different outcomes.

Gordon et al. (2015), too, came to the conclusion that the common cognitive and neural pathways of music and language processing allow rhythm and melody to facilitate language learning. When learners listen to and sing along with English songs, they naturally practice the rhythmic and phonetic aspects of the language in an enjoyable and engaging manner (Ludke & Ferreira, 2013).

Pronunciation accuracy is a critical component of language proficiency as it affects both comprehension and communication (Derwing & Munro, 2015). Pronunciation difficulties are often cited as major obstacles by English language learners and can impede their ability to effectively convey their intended messages (Sabouri et al., 2016). Consequently, innovative methods to improve pronunciation are of paramount importance in language education.

In recent years, the number and capabilities of AI-powered tools that help improve English pronunciation skills have made great advancements. These tools are equipped with "sophisticated speech recognition algorithms, machine learning, and Natural Language Processing (NLP) techniques to provide learners with adaptive learning experiences, real-time assessments, and personalized feedback" (Gupta & Garg, 2023; Mohammadkarimi, 2024, p. 100). Despite this significant rise in the power of AI, most research works present in the literature study the remit of AI in relation with grammar, vocabulary, and the like, leaving AI potentials in pronunciation enhancement underexplored

(Mohammadkarimi, 2024, p. 99). Even less frequent are the studies on AI-generated songs and their comparison with human-crafted ones, in terms of their effects on language learning and pronunciation accuracy in particular.

The application of AI-generated songs to pronunciation instruction remains empirically underexplored, despite the proven benefits of music in language learning. While AI tools offer personalized feedback on phonetic features such as stress and intonation, their musical adaptations introduce unique advantages, including dynamic tempo adjustment and error-targeted repetition (Mohammadkarimi, 2024). However, no controlled studies have directly compared AI-generated songs with human-composed ones, leaving unanswered questions about whether algorithmically created music can replicate the natural prosodic and emotional qualities that facilitate perceptual learning (Besson et al., 2011).

Recent generative AI advances now permit the creation of phonetically rich, grammatically simplified songs that bridge coursebook limitations and authentic materials (Gupta & Garg, 2023). Nevertheless, researchers call for rigorous experimental designs measuring not only immediate pronunciation gains but also long-term transfer effects and learner perceptions (Levis, 2018).

The scarcity of empirical research directly comparing the effects of AI-generated and human-composed songs on pronunciation outcomes represent a gap in the current literature. Although previous studies have reported benefits of using songs in pronunciation instruction (Farmand & Pourgharib, 2013; Kusnandar & Nurhasanah, 2020) and highlighted the potential of AI-powered tools for personalized language learning (Vančová, 2023; Karakaş, 2023; Mohammadkarimi, 2024), limited research has directly compared AI-generated and human-composed songs in terms of their contribution to EFL learners' pronunciation accuracy. In particular, further empirical investigation is needed to determine whether AI-generated songs offer advantages over human-composed songs in supporting the development of pronunciation features, including stress and intonation. This gap leaves educators with limited empirical guidance when selecting between these resources for pronunciation instruction. Moreover, the rapid development of generative AI technologies underscores the need for research evaluating their pedagogical value in specific language-learning contexts (Kohnke, Zou, & Zhang, 2024). Therefore, the present study addresses this gap by comparing the effects of AI-generated and human-composed songs on EFL learners' pronunciation accuracy.

2.1 Pronunciation

Pronunciation constitutes the physical manifestation of language, encompassing the articulated sounds, stress patterns, rhythmic contours, and melodic intonations that transform written symbols into spoken communication. Far from being a mere cosmetic feature of language use, pronunciation serves as the primary vehicle through which speakers convey meaning, emotion, and social identity in real-time interaction (Candlin & Crichton, 2019). The complexity of pronunciation lies in its dual nature: it involves both the cognitive processing of sound systems and the physical motor skills required for articulatory production. For second language learners, mastering pronunciation presents unique challenges, as it requires not only acquiring new phonological categories but also restructuring existing speech motor patterns shaped by their first language (Derwing & Munro, 2015). These challenges are compounded by the fact that pronunciation operates at the intersection of multiple linguistic levels—phonetics, phonology, morphology, and even syntax—making it one of the most integrative yet demanding components of language proficiency (Celce-Murcia et al., 2010). Given its

foundational role in oral communication, understanding the nature of pronunciation and its pedagogical implications is essential for any investigation into language learning and teaching methodologies, particularly as research continues to demonstrate that instructional approaches significantly shape learners' pronunciation outcomes (Alghazo et al., 2023; Thomson & Derwing, 2015).

2.1.1 Significance of Pronunciation in Language Learning

Unlike what is generally perceived, pronunciation is not a peripheral aspect of speaking, but a “pervasive feature of communication” shaping the “foundation of messaging” (Candlin & Crichton, 2019, p.1). Correct pronunciation also finds its importance in broader aspects of human life, such as identity and community membership(s). Identity is in turn connected with one’s social class, profession, education, place of birth, etc., which are altogether indicated by how words and sentences are pronounced. As Candlin and Crichton (2019) state, the significance of pronunciation is not limited to intelligibility, but extends to the broader sense of speaking and interpreting the speaker’s nature and orientation.

The positive effects of pronunciation instruction have been explored in the literature. For instance, Alghazo, Jarrah, and Al Salem (2023) refer to different research works, like that of Barrera-Pardo (2004) in which he reviewed 25 studies and reached the conclusion that 92% of the studies showed a positive effect of PI. In a more recent study, Thomson and Derwing (2015) conducted a narrative review of studies on PI and concluded that “[e]ighty-2% of the studies reported significant improvement” (p. 332). Alghazo et al. (2023) took a comparative approach and explored the efficacy of the type of instruction (perception-based vs. production-based) and found that each type may be proper for specific element of pronunciation. Together, these studies highlight the importance of instructional approaches in shaping learners’ pronunciation outcomes, which makes it essential to examine more closely what is meant by pronunciation accuracy and how it is defined in the literature.

2.1.2 Pronunciation Accuracy

Pronunciation accuracy has long been regarded as a central component of oral proficiency in second language acquisition. While the term often refers to the correct articulation of segmental features such as consonants and vowels, many scholars highlight the importance of suprasegmental elements—stress, rhythm, and intonation—in shaping accurate and comprehensible speech (Goodwin, 2014). However, whether to focus on intelligibility or nativelikeness has been a controversy in the field, affecting the meaning of accuracy. In this study, pronunciation accuracy is approached as the extent to which learners can produce segmental and suprasegmental features in a way that enhances intelligibility and reduces errors, making their speech clearer and more comprehensible.

In the investigation of the dichotomy of intelligibility versus nativism, there are various terms in the remit of pronunciation instruction that explain the qualities of pronunciation, such as intelligibility, accuracy, comprehensibility, interpretability, fluency, etc. As definitions of these terms may be arbitrary or overlapping, the research goal should have been clearly defined to mark its connections with the current trends of the field. Since the main focus of the present research was pronunciation accuracy, the researcher had to study the literature to fully comprehend where this goal stands, and if it had enough significance in the current trends. One such trend is the dichotomy of intelligibility and nativelikeness, in which according to Levis and LeVelle (2011) intelligibility wins.

The separation of accent and intelligibility is not a recent discussion. As Derwing and Munro (2015) state, Henry Sweet (1900) was probably the first to draw a line between the two. Intelligibility is concerned with the word and utterance level of recognition” and is considered the degree to which a message is heard as intended (Yang, 2020). They also note that comprehensibility and interpretability are above word level of meaning, and signal “metamessages” (Candlin & Crichton, 2019, p. 46). Munro et al. (2006) define comprehensibility as “the listener’s estimation of difficulty in understanding an utterance” (p. 112).

In their 2011 second language learning and teaching conference, Levis and LeVelle introduced the current goal of pronunciation teaching to be “helping learners to achieve comfortably intelligible pronunciation” (Yang, 2020, pp. 19-20). What is meant by an intelligible utterance is that “listeners can understand the speaker’s intended message” (Derwing and Munro, 2015, p.1). Therefore, lack of this quality leads to miscommunication. As a result, scholars such as Derwing and Munro (2015) have separated intelligibility from accentedness, indicating it as the main criterion of accurate pronunciation.

One reason for the recent unpopularity of nativism is that “few adult learners ever achieve native-like pronunciation in the L2” (Derwing and Munro, 2015, pp.6–7). In his study on automatic raters’ reliability in evaluation of Korean students’ pronunciation, Yang (2020) likewise noted that adult Korean learners of English now prefer acquire intelligible pronunciation, since nativelikeness seems unachievable to them.

2.2 Songs in Language Learning

From alphabet songs designed for children to popular songs by famous pop singers, it has always been witnessed that songs have facilitated language learning. Therefore, these life-long companies of human beings have been widely acknowledged in the literature to be one of the most powerful tools for their role in language learning, and more specifically, pronunciation improvement. Studies have shown that repeated exposure to songs enhances phonetic awareness and articulation skills.

Fonesca Mora (2000) studies the importance of music in any language classroom. She relies on Gardner’s (1993) theory of multiple intelligences and states that as there are different types of learning and learners, music should be included in language teaching. Shen (2009), too, confirms the need to incorporate music as it activates a valid type of learning and intelligence. Fonesca Mora (2000) and Gordon et al. (2015) also believe that the many shared features between language and music make their company ideal for learning a language. Their common features include their neural pathways, their originality being sounds, their leaning through exposure, and intrinsic features, such as volume, stress, intonation, rhythm, etc. Also, the ability to process sound-learning and auditory memory is among the very first abilities that develop in a fetus by hearing its mother’s heartbeat and sounds from outside the uterus. Therefore, even very young children are sensitive to the rhythmic features and musicality of their mother tongue, as expected (Fonesca Mora, 2000). Foneca Mora (2000, p. 147) believes that “discourse intonation, the ordering of pitched sounds made by a human voice, is the first thing we learn when we are acquiring a language”. What she refers to as ‘music’ in language teaching, is mostly relied on exaggerated intonations and singing through teaching, which is very similar to what mothers do to help their child learn better.

Shen (2009) supports the use of songs in English classrooms by bringing theoretical and practical underpinnings. She indicates that it is not only the melodic element of songs that make them suitable for teaching, but also the lyrics. In her opinion, lyrics “serve as a direct genuine source of teaching materials in foreign language classes”, and therefore should not be overlooked by teachers (p. 88).

Music, this popular universal language, has accompanied human beings and their shaping of language through history. Shen (2009) believes that history evidences how humans have used songs for their amusement as these melodic words have existed quite possibly as long or even longer than human speech. Many studies, too, now evidence the significant role that songs play in enhancing L2 pronunciation.

Historically, however, songs have been viewed more “as a form of entertainment rather than a medium of learning”. Nevertheless, integrating songs into language classroom activities is by no means a novel approach (Chan, 2022, p. 14). Richards (1969) highlighted the pedagogical value of songs in teaching vocabulary, rhythm, sentence patterns, and pronunciation of speech sounds.

The multifaceted benefits of songs in language learning naturally extend to one of the most crucial yet challenging areas: pronunciation. By exposing learners to authentic language through rhythm, melody, and repetitive patterns, songs provide a holistic input that supports both phonological processing and verbal output (Ludke, 2009). However, while the literature has extensively examined the contributions of songs to vocabulary retention and grammar acquisition, their role in enhancing pronunciation was underexplored until 2010s (Alghazo, Jarrah & Nour Al-Salem, 2023).

From 2010 onwards, the attention of some scholars shifted from vocabulary and retention to pronunciation improvement. In 2010, Salcedo explored the effects of songs on text recall and involuntary mental rehearsal. The study found that learners who engaged with songs demonstrated better memory retention, which in turn supported pronunciation accuracy, suggesting that songs function as effective mnemonic devices. Besson et al. (2011) highlighted that musical training can enhance linguistic abilities, particularly pronunciation, by improving attention and memory. Farmand and Pourgharib (2013) demonstrated that learners who frequently listened to English songs showed significant improvement in pronunciation skills. This study provides compelling evidence for the effectiveness of songs as supplementary tools in language instruction.

Ludke et al. (2013) demonstrated that singing can facilitate foreign language learning by promoting more natural pronunciation patterns and enhancing memory retention. Their study underscores the pedagogical benefits of integrating music into language instruction. Dewi et al. (2020) found that students who regularly listened to English songs exhibited significant improvements in their pronunciation abilities. The rhythmic and melodic elements of song lyrics help learners internalize pronunciation patterns and enhance their phonetic skills. The study highlighted the importance of incorporating dynamic and varied instructional techniques to optimize pronunciation training in language learning curricula.

2.3 Technology and Artificial Intelligence in Language Learning

In the context of language learning, technology’s application to pronunciation training represents a field of both promise and controversy. Although the wide impact of technology is

recurrently acknowledged by scholars, not unanimously do they agree on its sufficiency. An instance would be Derwing, Munro, and Carbonaro who - based on their research in 2000 - reached the conclusion that the advantages of using ASR in the procedure of giving feedback in PI was little to none. They believed that the humans were more accurate in recognizing the participants' utterances and their possible errors. "In other words, they were less adversely affected by L2 speech than was the software" (Derwing & Munro, 2015, p. 128).

Also, the cultural biases that may be inherent in AI systems, their insensitivity to cultural subtleties, inefficacy in natural language processing, etc. are among the disadvantages of AI-based language learning tools (Karakaş, 2023). Another challenge arises from the limitations of natural language processing (NLP) in handling diverse languages and dialects, including different varieties of English (Karakaş, 2023).

While the depth of its influence may seem unnecessary, intimidating, or insufficient to some scholars, the general view towards AI is quite welcoming. Of the most admired features of AI in the literature are its adaptability and accessibility for different learners. AI is considered to provide Language teachers and learners with opportunities that are "customized" and "catered to their unique requirements and preferences" (Karakaş, 2023, p. 216).

It is generally posited that one main challenge with conventional methods is in providing learners with tailored instruction and personalized feedback, where AI-driven pronunciation tools can appear as "catalysts for transformation" (Mohammadkarimi, 2024, pp. 98–99).

Songs are widely recognized as a useful resource for pronunciation practice, and artificial intelligence (AI) may further enhance their pedagogical potential through personalized and adaptable learning activities. Ludke (2009) recommends selecting songs that are not too fast and do not contain overly complex melodies or rhythms, as excessive difficulty may frustrate and demotivate learners. Similarly, Smith Salcedo (2002, as cited in Ludke, 2009, p. 10) suggests that songs containing too much unfamiliar vocabulary or grammar may confuse learners, whereas appropriately selected songs may facilitate verbal retention through their rhythmic and melodic patterns.

Computer-Assisted Language Learning (CALL) has expanded significantly in recent decades, with an especially large market in Asia (Candlin & Crichton, 2019). Similarly, Computer-Assisted Pronunciation Training (CAPT), applied in both language education and speech therapy, has seen rapid growth. As both areas continue to influence education and language learning, investigating their impact has become increasingly important. AI infinite affordances—covering accents, first-language backgrounds, and speech styles, multimodality, accommodation of repetition and imitation, etc. —far exceeds what learners are typically exposed to in a traditional classroom setting (Candlin & Crichton, 2019). Pronunciation as one of the most humane sub-skills of language, has found its connections with technology and AI through machine learning, natural language processing, speech recognition algorithms, etc., and its potentials are now being revealed more than any time before. AI-based pronunciation tools provide learners with greater flexibility and autonomy, enabling independent use anytime and anywhere. This accessibility broadens the reach of pronunciation instruction, allowing learners in varied educational contexts to benefit from such resources (Mohammadkarimi, 2024).

There is also evidence on the positive impacts of AI on pronunciation accuracy. Research indicates that well-designed CAPT resources can enhance pronunciation accuracy across multiple dimensions. Evidence also suggests that carefully designed CAPT systems can benefit children.

Despite these advantages, AI-driven pronunciation tools encounter several challenges. Candlin and Crichton (2019) note that current limitations in artificial intelligence, particularly in speech recognition and synthesis, prevent computers from fully engaging in authentic communication or meaning negotiation with humans. Whether these constraints can eventually be overcome, or whether they represent inherent boundaries of computational capability in comparison to human intelligence and language, remains uncertain.

Lee et al. (2021), in their meta-analysis on the effectiveness of pronunciation instruction, found that technology-based approaches generally produced smaller gains compared to human-led teaching. They suggest that this outcome may be attributed to computers' limited adaptability and perceptual accuracy relative to human teachers, which can, in turn, restrict their ability to provide appropriate feedback. Despite the promising affordances of AI-generated songs, several considerations must guide their implementation in pronunciation instruction. First, the quality of AI-generated musical content varies considerably, and poorly designed songs may inadvertently reinforce incorrect pronunciation patterns or fail to provide adequate models of target features (Candlin & Crichton, 2019). Second, the emotional and aesthetic dimensions of songs—factors that contribute significantly to their motivational impact—may be diminished when content is algorithmically generated, potentially reducing learner engagement over time. This concern echoes the broader critique that technologically sophisticated tools do not always translate into pedagogical effectiveness (Candlin & Crichton, 2019). Additionally, AI-generated songs, like other CAPT resources, should be evaluated against established principles of pronunciation pedagogy, including the emphasis on intelligibility over natelikeness and the importance of both perception and production practice (Derwing & Munro, 2015). Instructors are therefore advised to critically assess AI-generated materials, ensuring they align with learners' proficiency levels, address specific pronunciation needs, and complement rather than replace meaningful interaction and feedback. When used judiciously, AI-generated songs can serve as a valuable supplement to traditional instruction, offering learners engaging, accessible, and personalized opportunities for pronunciation practice (Mohammadkarimi, 2024).

3 Research Questions

This study seeks to investigate the effects of AI-generated and human-crafted songs on the pronunciation accuracy of EFL learners. To guide the investigation, specific research questions have been formulated to focus the scope of inquiry and ensure systematic data collection:

- RQ1. To what extent, if any, do AI-generated songs affect the pronunciation accuracy of Iranian EFL learners compared to human-crafted songs?
- RQ2. Is there a significant difference in the pronunciation accuracy of EFL learners who are exposed to AI-generated vs. human-crafted songs?

4 Method

4.1 Research Design

The present study employed a pretest–posttest quasi-experimental comparison-group design to compare the effects of AI-generated and human-composed songs on EFL learners' pronunciation accuracy. The participants were 42 upper-intermediate Iranian EFL learners from two intact classes at

a language institute. One group ($n = 20$) received instruction incorporating AI-generated songs, while the other ($n = 22$) used human-composed songs. Pronunciation accuracy was assessed before and after the intervention.

A quasi-experimental design was adopted because the existing classes were maintained for practical and pedagogical reasons, rather than randomly assigning individual learners to conditions. ANCOVA was used to compare posttest scores while controlling for pretest performance. However, this adjustment could not eliminate all potential pre-existing differences between the groups, such as motivation or prior exposure to English songs.

Potential threats to internal validity included selection bias, maturation, history, and testing effects associated with repeated use of the same assessment passage. Consistent assessment procedures helped support comparability, although they could not eliminate these threats entirely. The study's small sample, use of intact classes, and single-institute setting also limit causal inference and generalizability. Accordingly, the findings should be interpreted as evidence of differences associated with the two instructional approaches rather than definitive proof of causation.

4.2 Participants

The participants of this study were Iranian EFL learners from Pardisan Language Institute in Tehran, Iran. Initially, 56 learners aged 17–25 were recruited from the institute's face-to-face group classes. To ensure homogeneity in general English proficiency, the Quick Oxford Placement Test (QOPT) was administered. Learners who scored significantly below or above the upper-intermediate range were excluded, leaving a final sample of 42 upper-intermediate learners, composed of 25 female students and 17 male students.

The participants were drawn from two intact classes, which were assigned to the study conditions. Intact groups were selected for practical and pedagogical reasons, as they reflected the natural classroom context and ensured consistent exposure to the instructional treatments. The human-song group (comparison group, $n = 22$) received their regular instruction with human-crafted songs integrated to assist pronunciation learning. In the meanwhile, the AI-song group (experimental group, $n = 20$) received their regular instruction based on their weekly schedules, except that AI-generated songs were included in the instructional hours. Both male and female learners were included, and efforts were made to distribute genders across groups as evenly as possible.

Prior to data collection, participants were informed about the purpose and procedures of the study and assured that their participation was voluntary. Written consent was obtained from all learners, and they were notified that they could withdraw at any stage without any negative consequences. Anonymity and confidentiality of the participants' information were also guaranteed throughout the study.

4.3 Instruments

4.3.1 Quick Oxford Placement Test

The paper-based version of the Quick Oxford Placement Test (QOPT) developed in 2001 was employed to assess learners' general proficiency level and ensure homogeneity across groups. The QOPT has been widely validated in EFL contexts and is considered a reliable measure of overall English proficiency. Previous studies have reported an acceptable reliability index for the OPT ($r =$

.85) and confirmed its construct validity (Wistner et al., 2009). Since the study targeted upper-intermediate learners, the OPT was administered to determine participants' proficiency levels, and only those falling within the upper-intermediate band were retained for the final sample.

The test consists of two sections: Part One includes 40 questions, and Part Two comprises 20 questions. All 60 questions are multiple-choice and required approximately 50 minutes to complete. Scores ranging from 37 to 47 were targeted in this study, which represent upper-intermediate level.

4.3.2 Pronunciation Pretest and Posttest

The Rainbow Passage (Fairbanks, 1960) was selected as the standardized reading text for the pronunciation pretest and posttest. This passage has been widely used in phonetic and pronunciation studies due to its balanced phonemic distribution, natural rhythm, and inclusion of most English phonemes (see e.g., Munro & Derwing, 1999). It is therefore considered suitable and valid for assessing both segmental and suprasegmental features of speech among EFL learners.

The passage was printed and handed out to the participants. Before the actual recording, participants were given two minutes to review the text to minimize reading anxiety and unfamiliarity with lexical items. All readings were recorded using a high-quality digital recorder under identical conditions to ensure consistency. The same passage was used for both pretest and posttest sessions, allowing direct comparison of participants' pronunciation performance within and across groups.

The recordings were independently evaluated by the researcher and a rater expert in the field of TEFL to ensure inter-rater reliability. The assessment focused on both segmental features (consonants and vowels) and suprasegmental features (stress, rhythm, and intonation). In addition, to avoid any misjudgments by the raters, PRAAT software was used to double-check stress patterns and prosodic features.

The recordings were evaluated using a 5-point pronunciation rating scale adapted from Derwing and Munro (2005), assessing five criteria: vowels, consonants, word stress, intonation patterns, and rhythm. Each recording was independently rated by the researcher and another TEFL expert with experience in pronunciation teaching. Inter-rater reliability was established which yielded an acceptable level of agreement ($r = .87$).

4.3.3 PRAAT Software

PRAAT, a software tool for phonetic analysis, was used to supplement human evaluation of learners' pronunciation, particularly suprasegmental features related to stress. The analysis focused on three acoustic correlates of stress: fundamental frequency (F0), intensity, and duration. These measures can provide acoustic evidence relevant to stress placement, although none independently determines stress. PRAAT's pitch tracks, intensity displays, and spectrograms supported the examination of learners' speech productions and helped identify potential stress-placement errors that might not be readily apparent through auditory evaluation alone (Derwing & Munro, 2015; Thomson & Derwing, 2015). Thus, acoustic analysis complemented perceptual ratings and provided additional evidence for evaluating pronunciation accuracy. However, its contribution to assessment reliability depended on the consistency of the measurement procedures and the interpretation of the acoustic data.

4.3.4 AI-Generated and Human-Crafted Songs

The AI-generated songs were produced using SUNO software, with prompts designed and

customized by the researcher. Customizations were based on the content of each lesson and tailored according to factors such as topic, difficulty level, key vocabulary, length, tempo, genre, and voice type (male or female).

The human-crafted songs were selected based on their popularity, suitability for the learners' proficiency level, linguistic clarity, and cultural appropriateness. Both sets of songs were designed to closely resemble each other in terms of length, complexity, and overall instructional features.

Participants were exposed to the songs either as supplementary materials integrated into their regular lessons or as engaging activities during classroom sessions. In addition, learners were assigned out-of-class listening tasks for each session to strengthen their connection with the songs and increase their overall exposure.

4.4 Data Collection Procedure

Participants of the study were Iranian EFL learners drawn from two intact classes at Pardisan Language Institute in Tehran, Iran. Initially, 56 learners volunteered to participate in the quantitative phase of the study. To ensure homogeneity in general English proficiency, the Quick Oxford Placement Test (QOPT, 2001 version) was administered to the 56 volunteers. Learners outside the upper-intermediate band were excluded. The final sample comprised 42 upper-intermediate learners (25 females, 17 males), organized as the two original intact classes. The classes were then assigned to two conditions: Group A (experimental, $n = 20$) received their regular English instruction integrated with customized AI-generated songs, and Group B (comparison, $n = 22$) received their regular English instruction integrated with human-crafted songs. During the time of conducting the present research, the teachers of the two classes were supposed to teach units 5 and 6 of the third edition of American English File Upper-intermediate Student's book and Work book. The two units included various topics, such as anger, music, sleep, regrets, etc.

Following the intervention, the posttest replicated the pretest procedure: participants reread the same passage, recordings were rated independently by two raters, and PRAAT analyses were repeated to double-check the judgments of the prosodic features.

4.5 Data Analysis

The data were analyzed by a statistician expert in the field of TEFL using SPSS software. Descriptive statistics, including means and standard deviations, were calculated for pretest and posttest scores. Paired-sample t-tests were applied to examine within-group improvements in pronunciation, while independent-sample t-tests were used to assess between-group differences. Effect sizes (Cohen's d) were also calculated to determine the magnitude of observed improvements. To ensure rating consistency, inter-rater reliability was calculated using Cohen's Kappa coefficient.

5 Results

We used parametric analysis of independent samples t-test as four assumptions of interval data, independence of subjects, normality and homogeneity of variances were met. The first assumption is not violated because the present data are measured on an interval scale. Bachman (2005, p. 236) states that the assumption of independence of subjects is met when "the performance of any given individual is independent of the performance of other individuals" as this was true in the present study. The third assumption is the normality of the data which is

tested through the ratios of skewness and kurtosis. The results of normality test for OQPT scores are laid out in Table 1.

Table 1 indicates that the ratios of skewness (1.45) and kurtosis (1.01) over their respective standard errors for the OQPT scores of the AI-generated songs and the ratios of skewness (-1.37) and kurtosis (.38) over their respective standard errors for the OQPT scores of the Human-generated songs group are within the ranges of ± 1.96 denoting that the OQPT scores of the two groups did not violate the normality assumption. Also, Table 1 shows that the mean score of the AI-generated songs group ($M = 43.55$, $SD = 2.48$) and Human-generated songs group ($M = 42.86$, $SD = 2.39$) are close to each other.

Table 1

Descriptive Statistics of the Two Groups' OQPT Scores

Group	N	Mean	SD	SEM	Skewness Ratio	Kurtosis Ratio
AI	20	43.55	2.48	.55	1.45	1.01
Human	22	42.86	2.39	.51	-1.37	.38

According to the results manifested in Table 2, independent samples t-test detected no statistically significant difference in the English proficiency scores between the AI-generated songs and Human-generated songs groups, $t(40) = .91$, $p = .37$, $p > .05$ with the mean difference of just .69 on the side of AI-generated songs group.

Table 2

Independent Samples T-test for the Two Groups' OQPT Scores

Factor	Levene's Test for Variances		T-test for Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Diff.
Equal variances assumed	.02	.89	.91	40	.37	.69
Equal variances not assumed			.91	39.31	.37	.69

The results showed that regarding pronunciation accuracy pretest scores, the ratios of skewness (.79) and kurtosis (-.18) for the AI group, and the ratios of skewness (-.32) and kurtosis (-.20) for the human group over their respective standard errors are not beyond the ranges of ± 1.96 . It shows that these two sets of scores met normality assumption. Likewise, concerning pronunciation accuracy posttest scores, the ratios of skewness (-.57) and kurtosis (-.74) for the AI group, and the ratios of skewness (-.41) and kurtosis (-.74) for the human group over their respective standard errors are within the ranges of ± 1.96 denoting the normal distribution of the scores.

Note that the two raters scored pronunciation and the average of the two series of scores was used in the main analysis. As displayed in Table 3 and Figure 1, the mean of pronunciation accuracy in the AI-generated songs group ($M = 16.42$, $SD = 1.89$) and human-crafted songs group ($M = 15.75$, $SD = 1.85$) are close to each other on the pretest.

Table 3

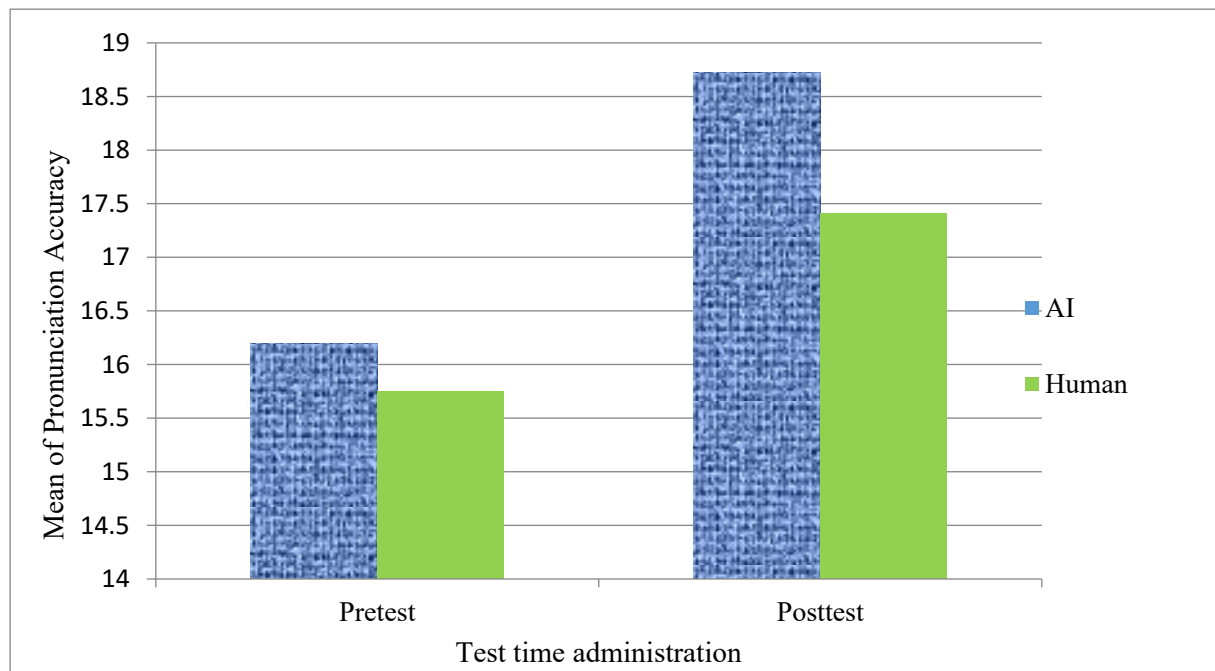
Descriptive Statistics of Pronunciation Accuracy Scores on Pretest and Posttest by Group (Two Raters' Average Scores)

Test	Group	N	Mean	SD	SEM
Pretest	AI	20	16.42	1.89	.42
	Human	22	15.75	1.85	.39
Posttest	AI	20	18.72	1.56	.35
	Human	22	17.41	2.01	.43

However, Table 3 and Figure 1 demonstrate that the mean of pronunciation accuracy in the AI-generated songs group ($M = 18.72$, $SD = 1.56$) is much higher than the mean in the human-crafted songs group ($M = 17.41$, $SD = 2.01$) on the posttest.

Figure 1

Bar Graph for Pronunciation Accuracy Means on Pretest and Posttest by Group



As represented in Table 4, paired samples t-test was significant for both AI-generated songs group, $t = 10.23$, $p = .00$, $p < .01$ with the mean difference of 2.30, and human-crafted songs group, $t = 7.33$, $p = .00$, $p < .01$ with the mean difference of 1.66. That means the learners of the two groups have progressed considerably from pretest to the posttest concerning pronunciation accuracy.

Table 4

Paired Sample Test for the Two AI-generated Songs and Human-Crafted Songs Groups

		Mean	SD	SEM	t	df	Sig.
Pair 1	AI	2.30	.98	.22	10.23	19	.00
Pair 2	Human	1.66	1.06	.23	7.33	21	.00

Paired-samples effect sizes were calculated using Cohen's (d_z), which standardizes the mean pretest–posttest difference by the standard deviation of the difference scores. For the AI-generated-song group, the effect size was ($d_z = 2.35$) (2.30/.98), whereas the human-composed-song group yielded ($d_z = 1.57$) (1.66/1.06; approximately 1.56 using the reported rounded value). These estimates indicate substantial within-group improvements in pronunciation accuracy over the intervention period. Both improvements were statistically significant: AI-generated songs, $t(19) = 10.23$, $p < .001$; human-composed songs, $t(21) = 7.33$, $p < .001$.

Although the standardized change was larger in the AI-generated-song group, these within-group effect sizes do not, by themselves, establish that AI-generated songs were significantly more effective than human-composed songs. The between-group comparison provides more direct evidence for that conclusion. In the present study, ANCOVA controlling for pretest scores indicated a statistically significant difference favoring the AI-generated-song group, $F(1, 39) = 6.13$, ($p = .01$), partial $\eta^2 = .15$.

The large within-group effect sizes should nevertheless be interpreted cautiously. They may partly reflect the relationship between pretest and posttest scores, the characteristics of the pronunciation measure, and the instructional context. Furthermore, the use of intact classes, the relatively small sample, and the absence of a delayed posttest limit the extent to which the gains can be attributed solely to the song type or assumed to persist over time.

Table 5

Effect Size for the Effectiveness of AI-Generated Songs and Human-Crafted Songs on Pronunciation

Group	Method	Effect size
AI	Cohen's d	2.35
Human	Cohen's d	1.56

ANCOVA assumes that the following assumptions are met: no influence of treatment on covariate measurement, reliability of covariates, no strong correlations among covariates, linear relationship between dependent variable and covariate, and homogeneity of regression slopes. Since the covariates were measured prior to the treatment, they could not be influenced by the treatment. Therefore, this assumption was not violated. In addition, there was only one covariate in each ANCOVA analysis. Hence, the assumption of correlation among covariates was not applicable.

To check the assumption of the reliability of covariates, inter-rater reliability was checked. Results showed that the covariate was measured reliably ($r = .83$). In order to do the analysis, the assumptions of linearity of the relationship between dependent variable and the covariate, and the homogeneity of regression slopes were also checked.

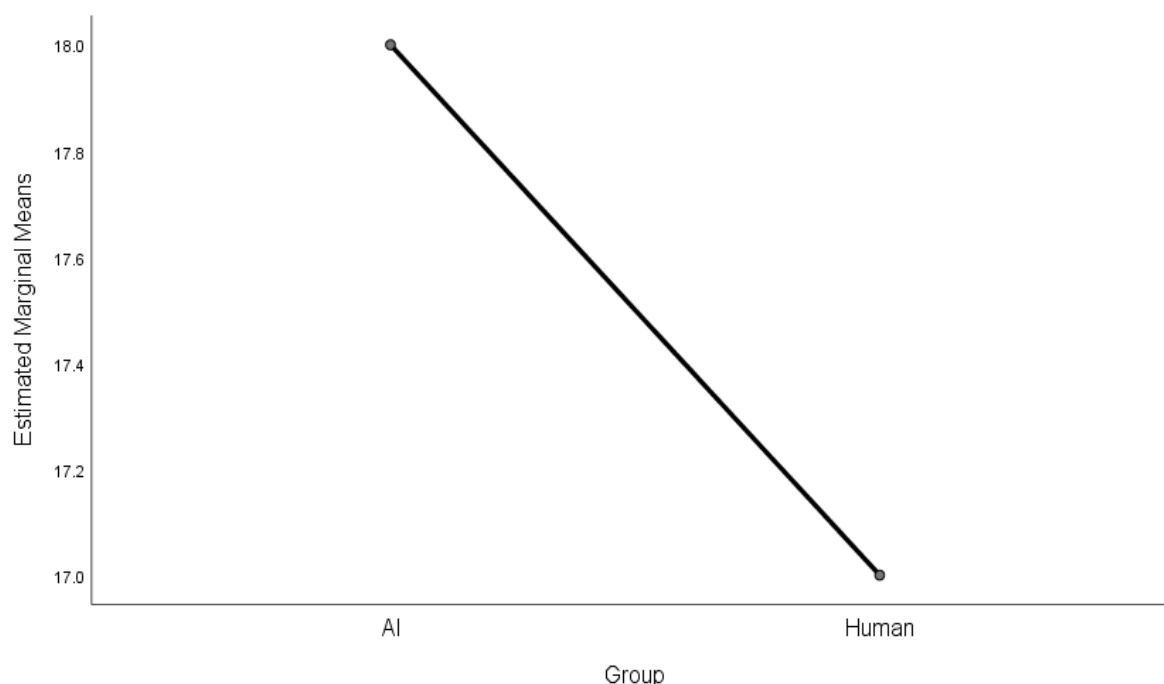
The estimated marginal means for pronunciation accuracy, adjusted for pretest scores, are demonstrated in Table 6 and Figure 2. With the pretest score held constant at 16.07, the AI-generated-song group obtained a higher adjusted posttest mean ($(M = 18.438)$, ($SE = .222$), 95% CI [17.988, 18.888]) than the human-composed-song group ($(M = 17.670)$, ($SE = .212$), 95% CI [17.242, 18.099]).

Table 6*Estimated Marginal Means for the Two AI-generated Songs and Human-Crafted Songs Groups*

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
AI	18.44 ^a	.22	17.99	18.89
Human	17.67 ^a	.21	17.24	18.10

a. Covariates appearing in the model are evaluated at the following values: Pretest = 16.07.

The adjusted mean difference was .77 points in favor of the AI-generated-song group. The confidence intervals indicate the estimated ranges for the adjusted means; however, the statistical significance of the between-group difference should be determined from the relevant ANCOVA test rather than from these intervals alone.

Figure 2*Line Chart of Estimated Marginal Means for the Two AI-generated Songs and Human-Crafted Songs Groups*

Covariates appearing in the model are evaluated at the following values: PRETEST = 16,071

The results revealed that the linear relationship between posttest of pronunciation accuracy and the covariate of pronunciation accuracy was significant, $F = 82.72$, $p = .00$, $p < .01$, therefore the linearity assumption was met. Also, the results of Levene's Test, $F(1, 40) = 1.07$, $p = .31$ indicated the homogeneity of variance assumption was not violated. The third assumption relates to homogeneity of regression slopes. The results, $F(1, 38) = 1.61$, $p = .21$ indicated that the assumption of homogeneity of regression slopes was met.

As presented in Table 7, after adjusting for the pronunciation accuracy scores on the pretest, there was a significant difference between the two groups' pronunciation accuracy scores on the

posttest, $F(1, 39) = 6.13$, $p = .01$, $p < .05$, partial eta squared = .15; thus, it can be claimed that using AI-generated songs is more effective than human-crafted songs to boost EFL learners' pronunciation accuracy. Additionally, as evident from Table 7, a strong relationship was seen between the pre-intervention and post-intervention scores on the total pronunciation accuracy, as shown by a p value of .00, $F(1, 39) = 95.71$, $p < .01$. Additionally, Table 7 shows that the partial eta squared (effect size) value is .69.

Table 7

Tests of Between-Subjects Effects for Pronunciation Accuracy Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	111.25	2	55.6	57.18	.00	.75
Intercept	13.42	1	13.42	13.79	.001	.26
Pretest	93.11	1	93.11	95.71	.00	.69
Group	5.97	1	5.97	6.13	.01	.15
Error	37.94	39	.97			
Total	13811.25	42				
Corrected Total	149.20	41				

6 Discussion

The present study examined whether AI-generated songs and human-crafted songs were associated with improvements in Iranian EFL learners' pronunciation accuracy and whether the magnitude of improvement differed between the two instructional conditions. Two main findings emerged. First, pronunciation accuracy improved significantly from pretest to posttest in both groups. Second, the AI-generated-song group demonstrated greater gains than the human-crafted-song group, and the difference in adjusted posttest scores was statistically significant, $F(1,39) = 6.13$, $p = .01$, partial ($\eta^2 = .15$). Although the within-group findings indicate improvement under both conditions, the between-group finding is particularly important because it suggests that the instructional condition may be relevant to the magnitude of learners' pronunciation gains. Given the quasi-experimental design and the relatively small sample, however, this difference should not be interpreted as definitive evidence that AI-generated songs themselves caused the greater improvement.

6.1 Within-Group Improvement in Pronunciation Accuracy

The first finding was that learners in both groups improved significantly in pronunciation accuracy following instruction. The AI-generated-song group increased its mean score from 16.42 to 18.72, representing a gain of 2.30 points, while the human-crafted-song group increased its mean score from 15.75 to 17.41, representing a gain of 1.66 points. These results suggest that pronunciation practice incorporating songs was associated with improvement under both instructional conditions. However, the gains within each group do not, by themselves, establish that songs were the sole cause of the improvement, because other instructional experiences and possible practice effects cannot be ruled out.

One plausible explanation for the improvement in both groups is the potential contribution of musical rhythm and repetition to the perception and production of language sounds. Rhythm can help learners attend to the temporal organization of speech, including stress patterns and the relative timing

of syllables. Because speech and music share some rhythm-processing mechanisms, repeated engagement with songs may provide opportunities to notice and reproduce phonological patterns. Fiveash et al. (2022) discuss the shared mechanisms involved in processing speech and musical rhythm, offering a theoretical basis for considering rhythm-based activities in pronunciation instruction. Similarly, Zhang et al. (2024) examined embodied music training in relation to speech imitation and skills, providing relevant evidence for the potential relationship between musical training and aspects of speech performance.

The present findings are also consistent with research investigating the use of English songs in pronunciation learning, including Farmand and Pourgharib (2013) and Kusnandar and Nurhasanah (2022). Such studies provide a broader rationale for incorporating songs into language instruction. Nevertheless, the improvement observed in the current study should not be taken as evidence that all songs or all forms of music-based instruction are equally beneficial. Dees and Cooper (2025), for example, reported that a phonics-integrated music-rhythm intervention did not significantly improve certain language-related outcomes. Although their study addressed different outcomes and a different instructional context, it cautions against assuming that the presence of rhythm or music necessarily produces learning gains. The effectiveness of music-based activities may depend on the target skill, the characteristics of learners, and how the materials are incorporated into instruction.

The contribution of the present study is that it documents improvement under two distinct song-based conditions within the same broader instructional context. However, because there was no non-song control group, the results cannot establish whether songs were more beneficial than conventional pronunciation instruction without songs. Furthermore, the use of the same reading passage at pretest and posttest means that familiarity with the passage may have contributed to the observed gains. Accordingly, the within-group findings support the pedagogical potential of song-based pronunciation practice but should be interpreted within these methodological constraints.

6.2 Greater Gains in the AI-Generated-Song Group

The second finding, and the principal comparative result of the study, was that the AI-generated-song group demonstrated greater improvement in pronunciation accuracy than the human-crafted-song group. After controlling for pretest scores, the difference in posttest pronunciation accuracy between the groups was statistically significant, $F(1,39) = 6.13$, ($p = .01$), $\eta^2 = .15$. The effect size indicates a meaningful difference in adjusted scores within this sample, although it does not establish that the same advantage would occur in other settings or with different learners and materials. The finding is therefore best understood as evidence that the AI-generated-song condition was associated with greater pronunciation gains in this particular study.

One possible interpretation concerns the degree to which instructional materials can be tailored to specific learning objectives. In the present study, the AI-generated songs were produced using SUNO, with prompts customized to the content of the textbook units and factors such as topic, difficulty, vocabulary, length, tempo, genre, and voice type. This degree of customization may have helped align the songs with the instructional content and the learners' needs. From a pedagogical perspective, the potential value of AI-generated materials may therefore lie not simply in their artificial origin, but in the opportunity to design and adapt musical input for a particular learning purpose.

A second possible explanation concerns repeated exposure to selected language features. Song-based activities can present words and sound patterns within a recurring rhythmic and melodic structure. Such repetition may give learners additional opportunities to attend to and reproduce target forms. AI tools can facilitate the creation of materials that are tailored to a lesson or revised to emphasize selected vocabulary and pronunciation features. Vančová (2023) discusses the potential of AI-powered tools for pronunciation training, while Mohammadkarimi (2024) explores the use of AI in promoting English pronunciation skills. These studies provide relevant context for considering AI-assisted pronunciation practice, although they do not, on their own, establish that AI-generated songs are superior to human-crafted songs.

Amrate and Tsai (2025), in their systematic review of computer-assisted pronunciation training, also provide a broader basis for examining technology-mediated pronunciation practice. However, computer-assisted pronunciation training includes a range of tools and instructional approaches, and its findings cannot automatically be transferred to AI-generated musical materials. The present study extends this area of inquiry by directly comparing AI-generated songs with human-crafted songs in a classroom-based EFL context. Its distinctive contribution is the observed difference between the two conditions, rather than simply another demonstration that technology-mediated pronunciation activities may be useful.

This finding also complicates the assumption that conventional human-crafted materials necessarily provide the more suitable model for language learning. Under the conditions examined here, the AI-generated-song group achieved greater gains. Nevertheless, the study did not isolate the mechanisms responsible for this difference. Although consistency of delivery, repetition, and alignment between the songs and instructional content are plausible explanations, these factors were not independently manipulated or measured. It would therefore be premature to conclude that AI-generated voices are inherently clearer, more accurate, or more effective than human singing. The comparative advantage may have reflected the specific songs selected, the way the AI materials were customized, differences in the musical or linguistic characteristics of the materials, or other unmeasured factors.

The result should also be interpreted alongside evidence that cautions against prioritizing technology over pedagogy. Neri et al. (2002) discussed the importance of the relationship between technological design and pedagogical objectives in computer-assisted pronunciation training. Their work supports the broader principle that technological innovation must be evaluated in relation to the learning needs it is intended to serve. Similarly, Derwing and Munro (2015) emphasize the importance of considering pronunciation in relation to communicative outcomes. In the present study, the outcome was pronunciation accuracy as assessed through a reading task; intelligibility, comprehensibility, and communicative effectiveness were not directly established. Consequently, the greater gains in pronunciation scores should not be assumed to indicate an equivalent advantage in everyday spoken communication.

Several methodological considerations further qualify the between-group finding. The sample consisted of 42 upper-intermediate learners from one language institute, and the classes were intact rather than individually randomized. These features restrict generalizability and limit causal inference. In addition, the use of the same passage for pretest and posttest may have introduced practice effects, while the absence of a delayed posttest means that the persistence of the gains remains unknown. It is

also unclear whether the raters were blinded to participants' group membership and test occasion. These limitations do not invalidate the observed statistical difference, but they indicate why the result requires replication under more controlled and diverse conditions.

6.3 Theoretical and Pedagogical Implications

Taken together, the findings suggest that the potential contribution of AI-generated songs to pronunciation instruction should be understood through the interaction between the instructional material and the learning task, rather than through the assumption that AI-generated content is inherently superior. Both groups improved, which is consistent with the possibility that structured song-based practice can support pronunciation learning. The greater gains in the AI-generated-song group suggest that tailoring materials to specific instructional content may be a relevant factor to investigate in future research.

Theoretically, these findings are compatible with accounts that emphasize attention to phonological patterns, repeated exposure, and the temporal organization of speech. Musical rhythm may help organize learners' attention to stress and timing, while customized lyrics may create opportunities to encounter target vocabulary and pronunciation features in a meaningful context. However, because the study did not directly measure attention, perception, memory, or the acoustic characteristics of the songs as explanatory variables, these processes remain plausible interpretations rather than demonstrated mechanisms. Future studies could test them by systematically varying song tempo, rhythmic complexity, lexical familiarity, and the degree of customization.

In practical terms, teachers may consider using AI-generated songs as supplementary materials when their content and musical structures are appropriate for the learning objectives. Their potential advantage is the possibility of creating lesson-specific resources, not an established capacity to replace human-produced materials. Human-crafted songs may continue to offer cultural, expressive, and artistic qualities that are valuable in language education. A balanced approach would therefore involve evaluating both AI-generated and human-crafted songs according to their linguistic suitability, naturalness, relevance to the lesson, and contribution to the pronunciation features being taught.

The present study did not directly investigate learner motivation, engagement, or attitudes toward AI-generated songs. Accordingly, no conclusions can be drawn about whether the AI condition was more motivating or engaging than the human-crafted condition. Such possibilities warrant separate investigation using learner questionnaires, interviews, classroom observations, or other suitable measures. Research should also examine whether the observed advantage persists over longer interventions and whether it extends beyond reading-based pronunciation accuracy to spontaneous speech, intelligibility, and communicative effectiveness.

In summary, both instructional conditions were associated with significant improvements in pronunciation accuracy, while the AI-generated-song group showed greater gains and higher adjusted posttest scores. This comparative finding adds preliminary evidence to the emerging discussion of generative AI in language teaching. However, the results should be interpreted in relation to the study's design, sample, intervention, and outcome measure. Further research is needed to identify the factors underlying the difference and to determine whether the advantage can be replicated and sustained across learners, contexts, and measures of pronunciation performance.

7 Conclusions and Implications

The findings of the present study indicate that both AI-generated and human-crafted songs were associated with statistically significant improvements in Iranian EFL learners' pronunciation accuracy. Although both groups demonstrated progress from pretest to posttest, the AI-generated-song group showed greater improvement than the human-crafted-song group. The ANCOVA results also revealed a statistically significant difference in posttest pronunciation accuracy between the two groups after controlling for pretest scores, with an effect size of partial ($\eta^2 = .15$). These findings suggest that AI-generated songs may offer a promising supplementary resource for pronunciation instruction. However, given the quasi-experimental design, the findings should be interpreted as evidence of an association between the instructional condition and pronunciation gains rather than definitive proof that AI-generated songs caused greater improvement.

The findings have several pedagogical implications. EFL teachers may consider incorporating AI-generated songs into pronunciation activities to provide learners with additional opportunities to listen to, repeat, and practice target pronunciation features. The adaptability of AI-generated materials may also allow teachers to tailor song content, vocabulary, tempo, and difficulty to learners' instructional needs. Nevertheless, AI-generated songs should complement rather than replace human-crafted songs and other established pronunciation practices. Teachers and material developers should evaluate the linguistic accuracy, intelligibility, naturalness, and pedagogical suitability of AI-generated materials before using them in the classroom.

Several limitations should be considered when interpreting these findings. First, the quasi-experimental design involved two intact classes rather than individual random assignment, limiting the strength of causal conclusions. Second, the relatively small sample of 42 learners, comprising 20 participants in the AI-generated-song group and 22 in the human-crafted-song group, restricts the generalizability of the results. Third, the participants were drawn from a single language institute in Iran, and the intervention covered two textbook units, limiting the extent to which the findings can be generalized to other educational settings, learner populations, and longer instructional periods. Fourth, pronunciation performance was evaluated using a five-point rating scale focusing on segmental and suprasegmental features; therefore, the results do not necessarily demonstrate corresponding improvements in intelligibility or overall communicative effectiveness. In addition, it is unclear whether the raters were blinded to group membership and testing occasion. The use of the same reading passage for the pretest and posttest also means that practice effects cannot be ruled out. Finally, the absence of a delayed posttest prevents conclusions about the durability of the observed gains, and the possible influence of the novelty of AI-generated songs remains unexamined.

Future research should address these limitations by recruiting larger and more diverse samples from multiple educational settings and employing more rigorously controlled research designs where feasible. Studies involving longer interventions and delayed posttests could establish whether pronunciation gains are maintained over time. Researchers should also consider using equivalent but different passages at pretest and posttest, ensuring blinded assessment where possible, and including measures of intelligibility and communicative effectiveness alongside pronunciation accuracy. Finally, further investigation of learner perceptions, engagement, and the possible novelty effect of AI-generated songs would provide a more comprehensive understanding of their pedagogical value.

Overall, the present study provides preliminary evidence that the use of AI-generated songs was associated with greater gains in pronunciation accuracy than the use of human-crafted songs in the investigated context. Although these results are encouraging, further research is needed to determine the extent to which this advantage can be replicated across contexts, sustained over time, and translated into broader improvements in spoken communication.

Acknowledgment

We are grateful to the authors of the research articles and textbooks used in this study.

Authors' Contributions

All authors have conducted the study, collected data, analyzed and interpreted the data, and written up the manuscript.

Funding

The study did not receive any funding.

Conflict of Interests

The authors declare that there is no conflict of interest.

References

- Alghazo, S., Jarrah, M., & Al Salem, M. N. (2023). The Efficacy of the type of instruction on second language pronunciation acquisition. *Frontiers*.
- Amrate, M., & Tsai, P. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 22–42. <https://doi.org/10.1017/S0958344024000181>
- Bachman, L. F. (2005). *Statistical analysis for language assessment (2nd ed.)*. Cambridge University Press.
- Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *frontiers in Psychology*.
- Candlin, C. N., & Crichton, J. (2019). *English pronunciation teaching and research*. Palgrave Macmillan.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. Longman.
- Dees, L., & Cooper, P. (2025). Effects of a phonics-integrated music rhythm intervention on reading fluency and accuracy with children. *Frontiers in Human Neuroscience*, 1–9. <https://doi.org/10.3389/fnhum.2025.1636278>
- Derwing, T. M., & Munro, M. J. (2005, September). Second language accent and pronunciation

- teaching: A research-based approach. *TESOL QUARTERLY*, 39(3), pp. 379–397.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals evidence-based perspectives for L2 teaching and research*. John Benjamins.
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410.
- Dewi, S. K., Sukarni, S., & Tusino, T. (2020). The influence of students' habit of listening to English songs on their pronunciation ability. *Scripta: English Department Journal*, 7(2), 14–24. <https://doi.org/10.37729/scripta.v7i2.827>
- Farmand, Z., & Pourgharib, B. (2013). The effect of English songs on English learners' pronunciation. *International Journal of Basic Sciences & Applied Research*, 2(9), 840–845.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE.
- Fiveash, A., Bedoin, N., Gordon, R., & Tillmann, B. (2022). Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology*, 771–791.
- Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. <https://doi.org/10.1017/9781108886901>
- Goodwin, J. (2014). Teaching pronunciation. In M. Celce-Murcia, *Teaching English as a Second or Foreign Language* (pp. 136–152). National Geographic Learning.
- Gordon, R., M.Fehd, H., & D.McCandliss, B. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in psychology*.
- Gupta, A., & Garg, A. (2023). AI-powered pronunciation tools: A systematic review of features and efficacy. *Journal of Educational Technology Systems*, 51(3), 289–310.
- Karakaş, A. (2023). Breaking down barriers with artificial intelligence (AI): Cross-cultural communication in foreign language education. In G. Kartar, *Transforming the Language Teaching Experience in the Age of AI* (pp. 215–233).
- Kohnke, L., Zou, D., & Zhang, R. (2024). Exploring generative AI in language education: Opportunities and challenges. *Computer Assisted Language Learning*, 37 (3), 1–24. <https://doi.org/10.1080/09588221.2024.2308573>

- Kusnandar, A., & Nurhasanah, S. (2022). Listening to English songs as a means of improving students' ability in pronunciation. *Elang: An English Education Language Journal*, 7(1), 1–7. <file:///C:/Users/PISHGAMAN09126026205/Downloads/AK+LISTENING+TO+ENGLISH+SONGS+AS+A+MEANS+OF+IMPROVING+STUDENTS%E2%80%99+ABILITY+IN+PRONUNCIATION.pdf>
- Kuzmanić, M. (2009). Validity in qualitative research: Interview and the appearance of truth through dialogue. *Psihološka obzorja / Horizons of Psychology*, 18(2), 39–50
- Lee, Z. (2021). A Critical Analysis of the Major Factors of Pronunciation Accuracy. *Erudite Journal of Linguistics and Languages*, 3(1), 33–41.
- Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.
- Ludke, K. (2009). *Teaching foreign languages through songs*. IMHSD University of Edinburgh
- Ludke, K., & Ferreira, F. (2013). *Singing can facilitate foreign language learning*. Memory & Cognition.
- Mohammadkarimi, E. (2024). Exploring the use of artificial intelligence in promoting English language pronunciation skills. *LLT Journal A Journal on Language and Language Teaching*, 27(1), 98–115. <https://doi.org/10.24071/llt.v27i1.8151>
- Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441–467. <https://doi.org/10.1076/call.15.5.441.13473>
- Oxford Placement Test: What does it measure and how? (n.d.). *Oxford University Press*.
- Pinar, S. (2023). A principled-review of studies in computer-based pronunciation instruction from 2015 to 2020. *ELT Research Journal*, 77–94. <https://dergipark.org.tr/en/download/article-file/3218064>
- Richards, J. C. (1969). Songs in language learning. *TESOL Quarterly*, 3(2), 161–174. <https://doi.org/10.2307/3586103>
- Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Routledge.
- Sabouri, N. B., & Pourhosein Gilakjani, A. (2016). Why i-s English Pronunciation ignored by EFL teachers in their classes? *International Journal of English Linguistics*, 195–208.
- Salcedo, C. S. (2002). The effects of songs in the foreign language classroom on text recall and

- involuntary mental rehearsal. *Louisiana State University and Agricultural & Mechanical College*.
https://doi.org/10.31390/gradschool_dissertations.1458?utm_source=chatgpt.com
- Shen, C. (2009). Using English songs: An enjoyable and effective approach to ELT. *English Language Teaching*, 2(1), 88–94.
- Tejedor-García, C., Escudero-Mancebo, D., Escudero-Mancebo, E., González-Ferreras, C., & Cardeñoso-Payo, V. (2016). Measuring pronunciation improvement in users of CAPT tool TipTopTalk! *Show & Tell Contribution*.
- Vančová, H. (2023). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, 11(3), 12–24. <https://doi.org/10.2478/jolace-2023-0022>
- Wistner, B., Hideki, S., & Abe, M. (2009). An analysis of the Oxford Placement Test and the Michigan English Placement Test as L2 Proficiency Tests. *Hosei University Repository*, 33–44.
- Yang, B. (2020). An evaluation of Korean students' pronunciation of an English passage by a speech recognition application and two human raters. *Phonetics and Speech Sciences*, 12(4), 19–25. <https://doi.org/10.13064/KSSS.2020.12.4.019>
- Zhang, Y., Baills, F., & Prito, P. (2024). Embodied music training can help improve speech imitation and skills. *Language Teaching*, 1–23. <https://doi.org/10.1017/S0261444824000363>



© 2026 by the authors. Licensee Journal of English for Specific Purposes Praxis, Iran. This is an open access article under the Creative Commons Attribution Non-Commercial 4.0 International (CC BY-NC 4.0 license) (<http://creativecommons.org/licenses/by-nc/4.0/>).