Effectiveness of Technology Assisted Pronunciation Training Devices for ESOL Learners

Camille Anolin

Northern Arizona University

Article Info

Keywords:
Keyword 1; Technology
Keyword 2; Second Language
Keyword 3; Pronunciation
Keyword 4; Intelligibility

Article History:
Received : 14/03/2025
Revised : 31/05/2025
Accepted : 02/07/2025
Available Online: 30/07/2025

Abstract

Technology Assisted Language Learning (TALL) has been a major focus in English education research for decades with a large number of devices and applications being utilized to promote general English proficiency along with targeting specific skills like reading, writing, listening, and speaking. This study investigates uses of TALL in English language classrooms, specifically TALL related to pronunciation training in the form of Technology Assisted Pronunciation Training (TAPT). This paper reviews research on TAPT in English language classrooms, evaluating the effectiveness of commonly used tools in enhancing language learning outcomes. Potential TAPT articles were identified through the search databases of the EBSCO and ProQuest with the key terms such as TAPT, TALL, Computer Assisted Pronunciation Training (CAPT), Mobile Assisted Pronunciation Training (MAPT), English for Speakers of Other Languages (ESOL), and pronunciation training. This review analyzes 10 empirical studies published between 2014 and 2024, encompassing a total of over 1,000 participants, to evaluate the impact of technology-assisted tools on pronunciation, intelligibility, and perception in English language learning. The review findings show that all studies consistently report positive results regarding the effectiveness of TAPT devices in enhancing intelligibility, suggesting their potential to significantly improve L2 English learners' experience in English classrooms.

Introduction

Over the past two decades, the accessibility of TALL (Technology Assisted Language Learning) has expanded significantly through the widespread use of smartphones and computers. However, pronunciation training, particularly in improving intelligibility, remains an area of limited focus in many L2 English classrooms. Research shows that, despite its effectiveness, pronunciation training is often overlooked in communicative language programs, with minimal class time dedicated to it (Meisarah, 2020). Many

instructors lack the skills to teach pronunciation or view pronunciation training as unsuitable for their students (Nair et al., 2017; Gilakjani, 2017, as cited in Stoughton & Kang, 2024). Therefore, TAPT (Technology Assisted Pronunciation Training) devices, along with CAPT (Computer Assisted Pronunciation Training), MAPT (Mobile Assisted Pronunciation Training), and VRPT (Virtual Assisted Pronunciation Training) tools, offer an alternative approach to pronunciation instruction through their smart features, valuable feedback, and gamified learning processes. Research conducted by Fouz-Gonzalez (2015) in the field of CAPT demonstrates the numerous tools available to help learners enhance their perception and production of various segmental and suprasegmental pronunciation features. These technological devices can provide L2 English learners with opportunities to improve their pronunciation both inside and outside the classroom, effectively addressing their learning needs.

This synthesis paper examines the effectiveness of TAPT devices in enhancing pronunciation intelligibility and learners' perception of pronunciation learning through the use of these tools. It aims to address the following research questions: (1) What devices are used for TAPT? (2) Are these devices effective in improving intelligibility for ESOL Learners? If so,

how effective? and (3) What are ESOL learners' perceptions in using TAPT devices for improving intelligibility of their speech?

Method

Access provided by the NAU Cline Library to databases such as EBSCO, and ProQuest was utilized to identify relevant empirical studies conducted within 10 years (2014 to present). This time frame was chosen due to the recent development and increased accessibility of TAPT devices, particularly CAPT, MAPT, and VRPT tools, which are now available through both paid and free software applications. The search terms 'TALL', 'TAPT', 'CAPT', 'MAPT', 'ESOL', 'pronunciation', and 'intelligibility' were utilized to identify relevant studies. The initial search utilized these key terms produced a total of 244 results. This pool of studies was then evaluated based on the following inclusion criteria:

1) The paper reports on a data-driven empirical study examining the improvement in pronunciation of L2 English learners while using TAPT tools

OR

2) The paper reports on a data-driven empirical and a survey-based study examining the improvement in pronunciation and L2 English learners' perception of using TAPT tools.

After applying these criteria, a total of 12 studies remained in the synthesis pool. Two papers were excluded because their distribution across the specific categories was imbalanced, with seven papers in CAPT, four in MAPT, and only one in VRPT. The final synthesis includes five studies on CAPT, four studies on MAPT, and one study on VRPT. A table outlining features of these devices is provided in Appendix.

Findings and Discussion

Devices used for TAPT

To thoroughly assess the effectiveness of the devices utilized in the TAPT, it is necessary to first review detailed descriptions of all devices used in the 10 primary studies. All information is systematically categorized according to specific device type: CAPT, MAPT, and VRPT. This approach allows for a structured analysis of each category, facilitating a clearer understanding of their respective contributions to the overall outcomes.

CAPT Devices

With the surge in available information and rapid technological advancements, Computer-Assisted Language Learning (CALL), especially CAPT, has seen significant development in recent decades (Calvo Benzies, 2017). CAPT tools and devices designed for pronunciation training and language learning, both inside and outside the classroom, have continued to grow and evolve. This section will provide an overview of five CAPT devices that have been recently utilized in research aimed at improving pronunciation and intelligibility, highlighting their key features and functionalities.

EPSSML (https://www.usc.gal/multimlab/) is an e-learning platform developed based on Mayer's (2008, 2009) Cognitive Theory of Multimedia Learning (CTML), which was designed to teach English phonetics and phonology within an EFL context. EPSSML's content is accessible both online and offline and has been specifically designed for Spanish-speaking learners of English. The platform provides a Sound Bank tab (vowels, diphthongs, etc.), Exercises tab (written & audio), Audio Illustrations, Resources, Glossary, and News tab.

Alongside addressing various phonetic and phonological topics, EPSSML places special emphasis on phonetic transcription training and focuses on areas identified as particularly challenging for Spanish-speaking learners according to multiple L2 speech

models. In the Sound Bank tab, the IPA phonetic symbols representing the sounds of received pronunciation are displayed on a keyboard, allowing them to be contrasted with those of the Peninsular Spanish speaker's perspective. EPSSML was studied by Gonzales and Ferriero (2024) in an empirical evaluation, which examined its efficiency in improving pronunciation intelligibility.

Reading Assistant SRS, a Speech Recognition System (SRS) developed by the American Scientific Learning Company, utilizes language learning theory to enhance learners' pronunciation through imitation and reading aloud exercises. The system accommodates variations in pronunciation ability across genders and ages. Its intelligent automatic scoring ensures accurate pronunciation imitation and controlled reading. Errors such as mispronunciations, misreadings, and inaccuracies are highlighted in different colors. Students receive feedback on their speech rate, error rate, and the most frequently misread phrases, in addition to a comprehensive read-aloud assessment. An empirical study of Reading Assistant SRS was conducted by Li (2020) which investigated its efficiency in pronunciation training.

Pronunciation Coach, a software developed by icSpeech, is designed to teach learners pronunciation, from individual sounds to full sentences. It allows students to record their speech and compare it to a model example, using speech recognition to score their pronunciation. The pronunciation model displays three views: the first shows the movement of the lips, tongue, and teeth; the second provides details on how each sound is produced; and the third highlights where the tongue contacts the palate. A study by Ghounane and Rabahi (2021) examined the effectiveness of teaching pronunciation using the Pronunciation Coach software.

Tell Me More English (v10), a software developed by SmartLink Corporation is an Automatic Speech Recognition tool that demonstrated strong consistency with human error detection and pronunciation rating (Bajorek, 2017). The system offers the option to practice pronunciation without internet connectivity. The system featured two types of activities: "listen and repeat" and "listen and choose the correct answer." These activities are designed to stimulate output, with no right or wrong answers. A study by Amrate (2018) examined the effectiveness of CAPT (i.e. Tell Me More v10) for pronunciation practice.

GeCall, Game Embedded CALL (GeCall) is a system developed specifically for a study conducted by Young and Wang (2014) that explored potential game strategies with automatic speech recognition technology. It incorporates a self-developed ASR application

to assess speakers' English pronunciation, evaluating factors such as tone, speed, volume, and timbre. The system's learning materials focus on essential elementary school vocabulary from textbooks, balancing language skills with game challenges to enhance learning through play. The game offers a four-level challenge to scaffold learners as they acquire vocabulary and complete tasks at their own pace. In drill practice, students listen to native speaker sounds, record themselves, and receive a score and feedback on their pronunciation, with the option for repeated practice. In summary, a variety of devices exist under the category of CAPT. These devices, though not restricted to particular features, facilitate pronunciation training by providing instruction on phonetic awareness, incorporating repetition and imitation, scaffolding and gamifying content, and offering instant feedback through ASR technology.

MAPT Devices

Mobile technology has drawn researchers' and educators' attention to Mobile-Assisted Language Learning (MALL), which uses handheld devices to support language learning. MALL is characterized by its potential to make learning personalized, spontaneous, informal, and accessible anywhere (Miangah & Nezarat, 2012). With pronunciation training getting less focus in MALL (Burston, 2015; Sung et al., 2016), this section examines four MAPT devices recently used in research to enhance pronunciation and intelligibility, outlining their key features and functions.

English Fun Dubbing is an application developed by Hangzhou Feizhu Technology Co., Ltd. and can be downloaded from all major application stores (i.e. Google Play & Apple Store) for free. This application offers English pronunciation and speaking practice through short videos, ranging from a few seconds to several minutes, featuring films, animations, documentaries, and storybooks. Learners can imitate the original audio sentence by sentence, and each imitation is recorded. Upon completing the online dubbing exercise, their recordings are automatically generated, evaluated, and scored. A study conducted by Wei et. al (2022) examined the effects of online English film dubbing activities on the intelligibility of learner's pronunciation.

English File Pronunciation App (EFP) is an application developed by Oxford University Press and can be downloaded from all major app stores (i.e. Google Play & Apple Store). The EFP app features an interactive sound chart that illustrates English sounds. The chart uses the same phonetic symbols and visual aids found in the English File book series. Learners can listen to sample words and sentences, with target sounds highlighted in a

different color. Additionally, they can record their pronunciation and compare it to the model for further practice. Learners receive instant feedback on their answers, with a green checkmark for correct responses and a red cross for incorrect ones. A study by Fouz-Gonzales (2020) examined the potential of using the EFP app to improve pronunciation of Spanish-speaking learners on specific segmental features.

TFlat is an application developed by TFlat Group and can be downloaded from all major app stores (i.e. Google Play & Apple Store) for free. This application includes four units, each focused-on teaching high and back vowels, diphthongs, and consonants. It also comprises four main sections where students can (1) check word pronunciation, (2) practice pronouncing words, (3) review sound details, and (4) record their voices. A study by Effat and Hamed (2018) examined the impact of using TFlat on pronunciation improvement in EFL learners.

ASR (Automatic Speech Recognition) with Google Documents, Google Translate, and Papago are applications developed by Google and Naver Corporation and can be downloaded from all major app stores (i.e. Google Play & Apple Store) for free. Google Translate and Papago are translation tools, while Google Docs is an online word processor for creating, editing, and collaborating on documents. All three apps utilize or can integrate ASR technology. A study by Dillon and Wells (2023) explored the use of ASR to assess and improve pronunciation intelligibility through reading aloud controlled passages. In this study, students read aloud controlled passages while ASR technology recorded their speech and converted it to text, with the resulting transcription reflecting their accuracy.

In summary, a variety of devices exist under the category of MAPT. These devices, while not limited to specific features, enhance pronunciation training by improving accessibility and offering a range of supports, including phonetic awareness instruction, repetition and imitation, content scaffolding, the use of authentic texts like dialogues and literature, and instant feedback through ASR technology.

VRPT Device

Virtual reality is recognized as an emerging technology with distinctive features that enable users to experience environments otherwise inaccessible (Minocha & Tudor, 2017). The technological innovations like virtual reality (VR) makes it possible to create educational experiences for learners who are distant from native environments (Lin & Lan, 2015). This section highlights a study conducted by Alemi and Khatoony (2020) using VRAPT administered by a humanoid robot and speech evaluation using Speech ace browser.

Speech ace (https://www.speechace.com/), is a pronunciation and fluency assessment software developed by SpeechAce LLC. This software does not require download and installation, and can be run through a browser (e.g. Google Chrome, Safari, etc.). This browser evaluates pronunciation accuracy and offers feedback to help language learners improve. It can assess non-native pronunciation, whether in simple or complex structures, by comparing it to that of a native speaker. Learners can record and replay their voice to check for native-like pronunciation, and the percentage of native-like pronunciation can be generated as a part of the feedback.

Effectiveness of TAPT

The second research discusses the effectiveness of TAPT devices in ESOL classrooms. For decades, educators and linguists have recognized the impact of integrating TALL into language classrooms, as studies consistently show significant benefits for students. Pope and Golub (2000) suggest that by engaging in a variety of computer-based tasks, students develop essential cognitive, metacognitive, and emotional skills. With TAPT devices and software becoming increasingly accessible, they are likely to become a standard feature in classrooms in the future. This section examines the effectiveness of TAPT tools for improving learners' pronunciation and intelligibility through empirical studies, highlighting key findings. For clarity and organization, Table 1 presents the essential data, followed by an analysis of the results.

Table 1

Study, participant data, study duration, device used, and key outcomes

Study	Participants	Study Duration & Device	Key Outcomes
Gonzales and Ferriero (2024)	Spanish, 504 college students	24 months EPSSML	Positive changes in formation of phonetic competence and motivation. Limited assessment in pronunciation performance.
Li (2020)	Chinese, 116 college students	12 months Reading Assistant SRS	Significant improvement in pronunciation, English proficiency, motivation.
Ghounane and Rabahi (2021)	Algerian, 80 college students	26 weeks Pronunciation Coach	Improved pronunciation in certain sounds and vowels. Awareness on differences between English and Arabic articulation of sounds.
Amrate (2018)	Algerian, 18 adult EFL learners	Six weeks Tell Me More (v10)	Small improvements in production of prosody features (i.e. syllable stress, sentence stress, and intonation)
Young and Wang (2014)	Taiwanese, 52 elementary students	Eight weeks GeCall	Game-based practice achieved better speaking improvement than those with only drill practice. Game-based activity was motivation to speak more.
Wei et. al (2022)	Chinese, 50 college students	Eight weeks English Fun Dubbing	Showed significant improvement in pronunciation, intelligibility and comprehension. Improved confidence and interest in learning English.
Fouz-Gonzales (2020)	Spanish, 52 college students	Two weeks English File Pronunciation App (EFP)	Positive effects in targeted segmental features and perception on target features. Three times the improvement compared to traditional instruction.

Table 1. *Cont.*

Study	Participants	Study Duration	Key Outcomes
Effat and Hamed (2018)	Iranian, 46 EFL learners	Six weeks TFlat	Noticeably higher progression in pronunciation ability.
Dillon and Wells (2023)	Korean, 59 college students	Four weeks ASR (Automatic Speech Recognition) with Google Documents, Google Translate, and Papago	Greater gains in overall pronunciation accuracy. Improvement on common pronunciation errors: LR, J/CH/Z, S/SH, and Vowels. regression on: BPFV
Alemi and Khatoony (2020)	Iranian, 18 young EFL learners	10 (90 mins.) sessions SpeechAce	High improvement on pronunciation performance via Cohen's effect size value. Enhance pronunciation skill particularly the vowels /ɪ/ and /i/ and vowels /ʊ/ and /u/

Analysis of the Learning Outcomes

Research consistently showed that using TAPT devices enhances pronunciation performance, with improvements closely linked to students' increased motivation when learning through these interactive tools compared to traditional instruction methods. It is important to note that younger learners (age: 7-12) demonstrated greater improvements in pronunciation than older learners. Studies by Alemi & Khatoony (2020) and Young & Wang (2014) found significant progress in younger participants, which contrasts with only modest gains observed among adult EFL learners in Amrate's (2018) study.

All 10 studies demonstrated a consistent pattern of accelerated improvement in pronunciation, increased motivation, and enhanced intelligibility among learners. For

example, Wei et al.'s (2022) study demonstrated significant improvements in pronunciation, intelligibility, and comprehension. Using the English Fun Dubbing device, students engaged in speaking practice through short videos ranging from a few seconds to several minutes, featuring diverse content such as films, animations, documentaries, and storybooks. Learners imitated the original audio sentence by sentence, with each attempt recorded and instant feedback provided. These features collectively enhanced students' confidence and interest in learning English, leading to positive learning outcomes.

While the studies largely demonstrated positive outcomes in pronunciation production, it is important to recognize that, in some experimental groups (e.g. Fouz-Gonzales, 2020), learners had access to TAPT devices outside the classroom, allowing for additional practice beyond what was available to control groups. This extra access highlights a potential advantage for experimental groups and suggests that frequency of device usage may have influenced pronunciation performance outcomes. Therefore, caution is needed when concluding that TAPT devices are definitely more effective than traditional instruction for teaching pronunciation intelligibility.

ESOL learners' perceptions of TAPT devices

While technology-assisted instruction is widely used in classrooms, TAPT devices are still underutilized in language learning. Yang (2001) notes technology's potential to positively influence learners' perceptions, attitudes, and phonological accuracy. This section examines learners' perspectives on TAPT usage, emphasizing its future importance in language learning. Although all studies reviewed here assessed participants' pronunciation (e.g., production of words, pronunciation intelligibility, or comprehensibility), only five included assessments of learners' perceptions on key TAPT features (e.g., xxxx, xxxxx). For clarity and organization, Table 2 presents the essential data, followed by an analysis of the perceptions.

Table 2

Perceptions of ESOL Learners in usage of TAPT devices.

Study	TAPT device	Survey Type	Perception
Gonzales and Ferriero	EPSSML	5-point Likert	Positive. 10 questions with averaging five to four points for each response. Motivation towards device usage rated an average of four.

(2024)			
Li (2020)	Reading Assistant SRS	Open-ended questionnaire	Positive. 80% praised automatic scoring function & correction adjustment. 90% benefited from SRS assisted learning. 94% favored integration to traditional classes.
Young and Wang (2014)	GeCALL	5-point Likert & qualitative questionnaire	Positive. Helped with word memorization, pronunciation, and speaking practice, and many were motivated to keep using it for further improvement. Rated to an average of four on all questions.
Wei et. al (2022)	English Fun Dubbing	5-point Likert	Positive. Learners citing its feasibility, effectiveness, and enjoyment; over half expressed a desire to repeat it multiple times weekly. Rated to an average of four on all questions.

Table 2. *Cont.*

Study	TAPT device	Survey Type	Perception
Amrate (2018)	Tell Me More English (v10)	Interview and Learning logs	Positive. Note that the program's UI is easy to use, activities are simple, and visual feedback is clear. However, while the collaborative CAPT group enjoyed the activities, the individual CAPT group found them monotonous.

Analysis of learner's perceptions

Results indicate that learners showed strong enthusiasm for using TAPT devices in pronunciation training and language learning. Surveys and questionnaires overwhelmingly recorded positive responses, with learners consistently expressing motivation to continue using the devices in future classes. Across all five studies, learners reported that blending TAPT with classroom lessons provided an engaging experience, with features like automatic feedback and scoring seen as particularly beneficial for their progress. For example, Li (2020) study questionnaire returned with an overwhelmingly positive result of 90% of

respondents reported being benefited from SRS-assisted learning, 94% favored integration to traditional classes, and 80% praised automatic scoring function and correction adjustment.

Young and Wang's (2014) qualitative data highlights that some learners felt less anxious about speaking, viewing the activities as a low-stakes game. Normally some students felt shy and feared being ridiculed by their classmates when speaking in front of those with better performance. However, this study observed that the speaking game encouraged greater interaction among students, allowing them to practice the target language together in a comfortable setting. Although students were moving around the classroom, the environment remained manageable, as they had to focus on the learning content to successfully complete the challenging barrier game.

In contrast, Amrate's (2018) study noted that some participants found certain TAPT activities monotonous, especially when compared to collaborative, peer-interactive sessions. This study formed two experimental groups: a collaborative group, where participants used the device with their peers, and an individual group, where participants worked alone. The content was divided into two parts. In the "listen and repeat" activities, participants listened to and repeated sentence models on a given theme (e.g., sports, travel, food). In the "listen and choose the correct answer" activities, participants listened to a short recording on a specific topic and selected the most appropriate response from several options. The second part, which involved selecting the correct answer, provided a valuable opportunity for collaboration among peers. This suggests that TAPT should be integrated into a blended approach to language learning, rather than used as a standalone, technology-focused solution.

Conclusions

This study explored the use of 10 TAPT devices and evaluated the effectiveness of these tools in assessing the impact of technology-assisted methods on pronunciation, intelligibility, and perception in English language learning within the classroom setting. The research concludes that the 10 studies reviewed offer strong evidence supporting the effectiveness of TAPT devices in enhancing learners' pronunciation intelligibility. TAPT devices improve pronunciation performance, with these advancements strongly tied to students' heightened motivation when using interactive tools compared to traditional teaching methods. All 10 devices used in the experimental groups consistently delivered better learning outcomes than the control groups. Only half of the studies assessed learners' perceptions, but they consistently reported that learners were motivated to continue using

TAPT devices in the classroom. Participants found the content more engaging and enjoyable, and viewed the automatic feedback as beneficial to their overall progress.

The findings suggest that fully integrating TAPT devices into the classroom can significantly benefit both learners and teachers by enhancing language learning outcomes. For teachers, these devices provide valuable support in focusing on improving students' pronunciation through interactive, engaging content and instant feedback. Beyond the classroom, TAPT devices promote self-learning by encouraging students to take initiative and ownership of their learning, helping them apply knowledge and develop new skills. Future research should investigate how current teachers can effectively adapt to this technology and their perceptions of its use in the classroom. Such exploration could enhance confidence among L2 instructors, learners, and researchers in the application of TAPT devices for pronunciation training.

REFERENCES

Ahmad, J. (2016). Technology assisted language learning is a silver bullet for enhancing language competence and performance: A Case Study. *International Journal of Applied Linguistics & English Literature*, 1-15.

Alemi, M. &. (2020). Virtual Reality Assisted Pronunciation Training (VRAPT) for Young EFL Learners. *Teaching English with Technology*, 59-81.

Amrate, M. (2018). A Mixed-method Approach to Exploring the Collaborative Practice of Prosody Features in Computer-assisted Pronunciation Training (CAPT): A Case Study of Algerian Undergraduate EFL Students.

Calvo Benzies, Y. J. (2017). Contribution of New Technologies to the Teaching of English Pronunciation. *Language Value*, 1-35.

ejedor-Garcia, C. E.-M.-A.-F.-P. (2020). Assessing Pronunciation Improvement in Students of English Using a Controlled Computer-assisted Pronunciation Tool. *IEEE*.

Fouz-González, J. (2020). Using Apps for Pronunciation Training: An Empirical Evaluation of the English File Pronunciation App. *Language Learning & Technology*, 62-85.

Ghounane, N. &. (2021). The Use of Computer Assisted Pronunciation Training in Teaching English Pronunciation for First-year EFL Students at Saida University. *International Journal of Applied Linguistics & English Literature*.

Gómez González, M. Á. (2024). Computer-assisted Pronunciation Training (CAPT): An Empirical Evaluation of EPSS Multimedia Lab. *Language Learning & Technology,*, 1-44.

Li, J. (2020). An Empirical Study on Reading Aloud and Learning English by the Use of the Reading Assistant SRS. *International Journal of Emerging Technologies in Learning (IJET)*, 103-117.

Lin, T. J. (2015). Language Learning in vVrtual Reality Environments: Past, Present, and Future. *Journal of Educational Technology & Society*, 45-70.

Meisarah, F. (2020). Mobile-assisted Pronunciation Training: the Google Play Pronunciation and Phonetics Application. *Script Journal Journal of Linguistic and English Teaching,*, 70-88.

Minocha, S. T. (2017). Affordances of Mobile Virtual Reality and Their Role in Learning and Teaching. *In The 31st British Human Computer Interaction Conference.* University of Sunderland's : University of Sunderland's St. Peter's Campus, UK.

Safavi, S. (2018). *Accent Conversion in Computer Assisted Pronunciation Training (CAPT).* Alliant International University ProQuest Dissertations & Theses.

Stoughton, A. &. (2024). A Systematic Review of Empirical Mobile-Assisted Pronunciation Studies Through a Perception–production Lens. *Languages* .

Wei, J. Y. (2022). Investigating the Effects of Online English Film Dubbing Activities on the Intelligibility and Comprehensibility of Chinese Students' English Pronunciation. *Academy Publication*.

Yang, S. C. (2001). Language Learning on the World Wide Web: An investigation of EFL Learners' Attitudes and Perceptions. *Journal of Educational Computing Research*, 155-181.

Young, S. S. (2014). The Game Embedded CALL System to Facilitate English Vocabulary Acquisition and Pronunciation. *Educational Technology & Society*, 239-251.