SALEE: Study of Applied Linguistics and English Education
Vol. 6, No. 2, July 2025

SALEE
Study of Applied Linguistics and English Education
P-ISSN: 2715-9795 E-ISSN: 2716-1617

Exploring Technology Use in English Pronunciation Instruction: A Systematic Review of Trends and Tools

Nursyafida [1]
Reni Eka Putri [2]

[1] Postgraduate Program of English Education, State Islamic University of Sultan Syarif Kasim Riau, Indonesia;
[2] Corresponding author, Postgraduate program of English Education, State Islamic University of Sultan Syarif Kasim Riau, Indonesia; reni68693@gmail.com

Received: 12 January 2025 *Accepted: 19 July 2025* *Published: 26 July 2025*

Abstract

The integration of technology into education has reshaped pronunciation instruction, introducing innovative tools like Computer-Assisted Pronunciation Training (CAPT) and speech recognition systems. This study aims to systematically review the trends and tools used in technology-assisted pronunciation learning and evaluate their effectiveness. Employing the PRISMA protocol, 27 peer-reviewed studies published between 2006 and 2024 were analyzed through a systematic review method. The findings reveal a significant rise in the use of Computer-Assisted Pronunciation Training (CAPT), mobile applications, and speech analysis software such as Praat, Audacity, and ASR-based systems. These tools offer interactive features, immediate feedback, and personalized learning, contributing to improved phonetic accuracy across segmental and suprasegmental aspects. The majority of studies focused on learners in higher education, ranging from beginner to advanced levels. While results demonstrate the positive impact of technology on pronunciation proficiency, limitations remain in long-term impact evidence and contextual adaptation, particularly in under-resourced settings. This study underscores the potential of digital tools to complement traditional methods and recommends future research to address sustainability, accessibility, and pedagogical integration.

Keywords: CAPT, digital tools, pronunciation instruction, systematic review, technology integration

To cite this article: Nursyafida & Putri, R. E. (2025). Exploring technology use in English pronunciation instruction: A systematic review of trends and tools. *SALEE: Study of Applied Linguistics and English Education*, *6*(2), 414-436. https://doi.org/10.35961/salee.v6i2.1931

DOI: 10.35961/salee.v6i2.1931

1. Introduction

The increasing use of technology in education in the current era of globalization, educators are increasingly aggressive in using technology in the learning and training process, where the use of technology is assumed to be one of the influences in student learning. This is in line with the opinion of Costley (2014) who stated that student learning is positively influenced by technology where technology makes students more involved in the learning process, and they tend to remember knowledge better. In addition, the use of technology is used every day to communicate, obtain information, so the use of technology is very relevant to students and provides connections that will be very beneficial for student learning.

Today, technology is an integral part of the learning experience and is an important consideration for teachers. The important role that technology plays in education provides teachers with the opportunity to design meaningful learning experiences that create technology (Eady & Lockyer, 2013). So, it can be said that technology plays a crucial role in the world of learning which brings various significant benefits to the learning process such as increasing students' interest and motivation in learning.

In language teaching, especially English language learning, the use of technology is also an important concern, one of which is teaching and learning pronunciation. According to Gilakjani (2016), pronunciation is one of the most difficult skills in the learning and teaching of English language. Naser (2022) also stated that there are several factors that influence pronunciation difficulties, namely the influence of the mother tongue and limited exposure to native pronunciation models. In addition, the main problems in learning English pronunciation are related to socio-psychological variables such as anxiety, self-confidence, motivation, attitudes, and so on. So, teaching pronunciation also really needs attention.

Pronunciation teaching is an important component of learning because proper pronunciation improves communication skills and comprehension. However, traditional teaching methods are often less effective in meeting the individual needs of students. As technology advances, various tools and platforms have been developed to support pronunciation teaching, such as Computer-Assisted Pronunciation Training (CAPT) and others (Amrate & Tsai, 2024). In teaching and learning pronunciation the use of CAPT technology is very often used, one of the studies conducted by Chu Liu and Po-Yi (2016) showed that the use of CAPT in teaching pronunciation has a significant increase in pronunciation quality, which findings provide empirical evidence that the use of computer technology is effective in teaching pronunciation.

Although various technologies have been applied in pronunciation teaching, there are still some gaps in the literature regarding what technologies can be used in pronunciation teaching and how effective they are, especially for learning in Indonesian schools. Therefore, this article aims to fill the gap by systematically reviewing trends and tools in technology-based pronunciation teaching, evaluating their effectiveness, and providing recommendations for future research and practice.

For this aim, this study is guided by the following research questions:

1. What technologies are used in pronunciation learning?
2. How effective is technology in improving English pronunciation skills among students?

The results of this study also intended to be a guide for teachers and educational administrators when selecting devices and for teaching pronunciation in schools. This study aims to assess the effectiveness of various technologies, such as applications used in improving students' pronunciation skills at various levels of education.

1.2. Theoretical framework

1.2.1. Technology in teaching

Technology refers to the tools, methods, or knowledge employed to accomplish specific tasks. In this context, technology is defined as the application of technical processes, methods, or knowledge to achieve learning objectives or instructional goals. For example, Merriam-Webster (2019) explains that the students can complete a speech activity by describing the features of an object in the target language while using a digital voice recorder to capture their speech. They can then share the recording with their teacher for feedback. The integration of classroom technology enhances students' ability to succeed in learning English. Beyond traditional teaching methods, technology provides numerous learning opportunities, offering teachers access to a wide array of resources and materials for their lessons. Cutter (2015) also highlights that technology increases student engagement. Students are more motivated when using computers and modern devices compared to traditional textbooks.

They are often less distracted and for each student the teacher can choose personal lessons according to their needs. In addition, students can use their home appliances to complement their knowledge. In the learning environment, technology has always been a major factor. Altun and Ahmad (2021) identified that technology has consistently played a significant role in educational settings, becoming a key element of the teaching profession to facilitate learning for students. The term integration is commonly associated with the inclusion of technology in teaching and learning. Since technology is an integral part of daily life, it is essential to rethink how it can be incorporated into curricula to enhance the teaching process. Integrating technology in education broadens teachers' abilities and provides numerous

opportunities to teach English effectively. For example, using English-language videos or songs related to the lesson topic can help students grasp concepts more quickly and develop a more positive perspective on the subject matter.

Furthermore, Murray (2005) explained that the integration of next-generation technology provides students with a natural context for developing learner autonomy, shaping their identity as learners, discovering new ways to use language, and fostering motivation to engage in collaborative interactions with teachers and peers. In addition, Gilakjani (2017) showed that the technology significantly enhances students' activities and exerts a profound influence on teaching methodologies. Teachers who fail to incorporate technology into their teaching risk falling behind in its advancements. Consequently, Solanki & Shyamlee (2012) stated that educators must develop comprehensive expertise in utilizing technology to teach language skills effectively.

Nomass emphasizes various technological tools that are particularly valuable for teaching speaking and listening skills. These include web-based language learning platforms, innovative presentation software, online dictionaries, chat features, email, and computer-assisted language learning programs. His research, conducted as a case study, explored the use of technology by a group of students learning English in a school setting. Nomass (2013) also addresses challenges in modern classrooms, particularly the gap between theoretical approaches and practical applications of technology. He advocates for the integration of technology into language learning practices and stresses the need for educators to embrace such innovations to improve classroom outcomes.

1.2.2. Integrating technology in the classroom

Using the internet in education significantly enhances student motivation. Altun and Ahmad (2021) explained that incorporating films into learning allows students to engage with and expand their knowledge more enthusiastically. Through computers and internet access, technology enables students to participate actively in their learning processes, facilitating more effective learning outcomes. Furthermore, Arifah (2014) learning with technology encourages the development of higher-order thinking skills. In summary, combining multimedia with appropriate teaching methodologies is crucial for capturing learners' interest in the English language.

Two approaches to integrating technology into the classroom have been highlighted, referring to Warschauer (2000). The first, the cognitive approach, emphasizes creating meaningful contexts for learning where students can maximize their language exposure and build individual knowledge. Examples include text-reconstruction software and multimedia simulation software. Multimedia simulation tools immerse learners in computerized micro-worlds with rich audiovisual contexts focused on language and culture. Advanced programs

in this category allow students to interact, manipulate, and control their linguistic input, enhancing their learning experience.

The second, the social approach, focuses on the social aspects of language learning, viewing it as a process of socialization. In this approach, students engage in real-life skill development through authentic social interactions. This is achieved through collaborative projects and tasks that mimic real-world scenarios (Warschauer, 2000).

A key element of language learning is pronunciation. However, it is difficult to provide helpful feedback. In traditional teaching methods, Zhao (2013) identified that the teachers often provide feedback and modeling based on their interpretation of the student's pronunciation, which may not always be accurate. Typical feedback involves asking students to repeat pronunciations or providing abstract explanations of how to produce sounds. However, advancements in speech recognition technology enable more precise and efficient feedback for learners.

The use of technology in classrooms in English as a foreign language can provide an important and attractive approach to linguistic study. Ilter (2009) explained that it motivates students and enhances their ability to speak, read, listen, and write effectively. However, technology is not enough to teach ELLs alone. Schwartz and Pollishuke (2013) stated that the teachers must set clear objectives, understand the curriculum, and employ effective teaching strategies to provide meaningful learning experiences that build upon students' prior knowledge.

A study by Lin and Yang (2011) have demonstrated that Wiki technology improves the written skills of students. The students participated in a Wiki project where they wrote passages, read their classmates' contributions, and provided responses. Learners reported that this method was beneficial because of the immediate feedback they received. Additionally, students improved their vocabulary, spelling, and sentence construction by reviewing their peers' work.

A study by Peregoy and Boyle (2012) showed that using technology to improve the reading and writing skills of learners. The results of this study showed that the tools used in technology enhance the ability of learners to write and read because they are user friendly. The other results of this study were that leaners learn more efficiently by using technology tools instead of traditional teaching methods because the Internet provided a conducive learning environment for learners and provided a new platform for students who can access learning lessons in a convenient way. Additionally, Rodinadze and Zarbazoia (2012) identified that the technology supports both students and teachers by enabling quick access to educational materials. Technological advancements also play a crucial role in preparing students for the global workforce, serving as practical tools that facilitate meaningful learning.

The use of computers and other technological devices should be encouraged, since this can increase the motivation of students in English courses and can achieve efficient and successful goals. However, Ilter (2009) emphasized that technology should not be relied upon as the sole source of motivation. Instead, lessons must be carefully planned with thoughtful use of technology and grounded in sound educational principles. Gustad (2014) also highlights that the use of technology in teaching has led to a marked increase in student motivation. Prior to the availability of internet access in schools, students were limited to the information found in classroom textbooks and school libraries. With open internet access, students now have powerful tools for conducting research and engaging in deeper learning experiences.

1.2.3. Pronunciation teaching and learning instruction

Interest in the use of technology to support language learning, particularly through Computer-Assisted Language Learning (CALL) and Computer-Assisted Pronunciation Training (CAPT), has grown significantly in recent decades. Pennington and Rogerson-Revell (2019) explain that both CALL and CAPT have had profound impacts on education, language learning, and teaching, making research on their effects crucial. While these technologies offer significant educational benefits, such as increased motivation through their novelty, access to multimodal resources (audio, print, and video), and the flexibility of anytime-anywhere learning, their commercialization often prioritizes sales over educational value. CAPT, in particular, offers numerous advantages, including endless opportunities for repetition and imitation, instant feedback, exposure to diverse target language speech, and individualized, self-paced learning. A key feature of CAPT is its ability to provide automated feedback, which is especially valuable for pronunciation practice. Moreover, technologies used for pronunciation instruction often have applications in research and assessment, further enhancing their potential.

Pennington and Rogerson-Revell (2019) identify that speech analysis enables the acoustic evaluation of speech signals, often visualized as waveforms, speech contours, or spectrograms. Tools such as Wavesurfer and Praat, widely used in phonological research, provide free access to these features, while many commercial CAPT programs integrate them. Automatic Speech Recognition (ASR), or speech-to-text (STT), translates spoken language into written words and is a key area of linguistic research and technological development. Examples include commercial software like Dragon Naturally Speaking. On the other hand, speech synthesis involves generating computer-produced speech. This can include text-to-speech (TTS) systems or advanced programs that simulate speech through articulatory movements or phonological rules. Combined ASR and speech synthesis systems create dialogue-based technologies used in speech training. Notable advancements in these fields have paved the way for intelligent personal assistants like Siri, Alexa, and Cortana, which can process diverse queries and respond naturally.

Developments in educational technology have led to a wider range of delivery mechanisms, from more traditional companion websites and CD versions of books, such as The Cambridge English Pronouncing Dictionary and Clear Speech (Gilbert, 2012), to mobile applications like Pronunciation Power and Sounds, as well as social media platforms such as Twitter, Facebook, and WhatsApp. Some products, like Pronunciation Power, offer multiple formats, including DVD, interactive websites, and mobile apps, making them highly versatile. According to Pennington (1999), these resources are particularly advantageous in overcoming limitations of traditional language classrooms. They have the potential to offer a private, stress-free environment where learners can access virtually unlimited input and practice at their own pace; and, with the integration of Automatic Speech Recognition (ASR), they can provide individualized, instantaneous feedback.

1.2.4. Student engagement in technology-based pronunciation instruction

Student engagement has become a central focus in language learning, particularly in environments that incorporate digital tools. Engagement refers not only to students' behavioral participation in tasks but also to their emotional investment and cognitive involvement in learning processes (Fredricks, Blumenfeld, & Paris, 2004). In the context of pronunciation learning, engagement is essential because it encourages sustained practice, repetition, and attention to phonetic detail that's key elements required to master segmental and suprasegmental features.

The integration of digital technologies such as CAPT (Computer-Assisted Pronunciation Training), mobile applications, and visual audio tools like Praat and Audacity has been shown to enhance student engagement by providing immediate feedback, interactive exercises, and visually stimulating representations of sound. According to Gilakjani (2017), the use of technological tools in pronunciation teaching significantly increases students' participation, concentration, and overall motivation. Interactive features such as real-time feedback, visual waveform display, and gamified repetition routines create a more dynamic and immersive learning experience.

Furthermore, technology offers a multisensory approach that caters to various learning styles, enabling students to engage in deeper, more meaningful interactions with pronunciation content. As stated by Cutter (2015), students demonstrate higher engagement when tasks are supported by digital tools that appeal to both auditory and visual processing. In pronunciation instruction, this is particularly beneficial since learners must simultaneously process sound quality, mouth positioning, rhythm, and stress that all of which can be demonstrated and analyzed more effectively through software.

CAPT platforms such as MyET or iSpraak simulate teacher-like feedback loops, allowing learners to receive phonetic correction in real-time. This feature fosters repeated engagement as students become curious to improve their scores or correct errors. Fryer and Carpenter

(2006) suggest that such immediate interaction with a system encourages behavioral engagement, as learners are more likely to remain on-task and motivated to refine their output through practice.

Moreover, gamified applications like ELSA Speak or YouGlish incorporate elements of challenge and progression, which sustain emotional engagement. These systems track user performance and provide scores, badges, or progress charts, factors that support Dörnyei's (2001) motivational framework by creating short-term goals within long-term learning. Emotional engagement is crucial in pronunciation learning, where frustration due to incomprehensibility or embarrassment is common.

Therefore, digital pronunciation tools not only support the technical aspects of phonetic training but also improve the quality of student engagement. They enable consistent participation, emotional investment, and sustained focus through adaptive and interactive environments. As engagement is a predictor of academic success (Reeve, 2012), its integration in pronunciation pedagogy via technology enhances both the learning experience and long-term pronunciation gains.

1.2.5. Learner autonomy and self-regulated pronunciation practice

Learner autonomy refers to the capacity of learners to take control of their own learning process, including setting goals, choosing strategies, and evaluating progress (Little, 1991). In pronunciation instruction, autonomy plays a critical role because mastery of pronunciation requires continuous and individualized practice that extends beyond the classroom. Technology provides a framework for autonomous learning by offering learners flexible, self-directed environments with minimal teacher supervision.

Technological tools such as CAPT programs, ASR-based apps, and mobile pronunciation platforms empower learners to practice pronunciation anytime and anywhere. As pointed out by Murray (2005), the integration of technology in language instruction provides authentic contexts for learners to exercise autonomy, shaping their identities as proactive language users. Learners using tools like Praat or Audacity are able to record, analyze, and compare their speech with native models, activities that reflect goal setting, self-monitoring, and metacognitive regulation.

Moreover, speech analysis tools enable learners to visualize their errors in pitch, duration, and articulation, thereby making abstract phonetic concepts more concrete. According to Gilakjani and Rahimy (2019), students who engage with CAPT systems demonstrate increased independence in identifying and correcting their own pronunciation errors. This aligns with Benson's (2011) view that learner autonomy is not merely about choice, but about the development of self-efficacy through structured, yet self-managed learning processes.

Applications such as Pronunciation Power 2, Google ASR, and EPT allow users to receive personalized feedback without the presence of a teacher. This kind of environment not only

supports self-paced learning but also encourages learners to take responsibility for their improvement. Chu Liu and Po-Yi Hung (2016) found that learners using CAPT tools showed higher levels of ownership and accountability for their progress compared to those in traditional instruction settings.

Importantly, autonomy is also linked to motivation and persistence—two critical factors in pronunciation learning. As pronunciation improvement is often slow and incremental, the ability to persist through self-driven practice is essential. Autonomous learners using technology can choose to repeat tasks, review past recordings, or customize exercises based on their weaknesses, which aligns with the principles of adaptive learning.

Thus, by placing learners at the center of decision-making and control, technology supports the development of autonomy in pronunciation instruction. It encourages reflection, strategic practice, and lifelong learning skills, all of which are necessary for meaningful pronunciation development in real-world contexts.

1.2.6. Higher-order thinking skills in digital pronunciation learning

Higher-order thinking skills (HOTS) include cognitive abilities such as analysis, synthesis, evaluation, and problem-solving (Bloom, 1956). In pronunciation instruction, HOTS are often overlooked in favor of mechanical repetition. However, the use of advanced technologies in pronunciation learning has the potential to engage learners in cognitive tasks that go beyond rote imitation.

When students use speech analysis tools like Praat or ASR systems, they are required to interpret acoustic patterns, compare outputs, and make decisions about how to improve specific phonetic features. This process involves critical listening, error analysis, and hypothesis testing, which align with the analytical dimension of HOTS. According to Zhao (2013), students using technologically enriched environments develop higher levels of phonological awareness and evaluative judgment.

In particular, technologies that provide spectrograms and visual representations of pitch and stress encourage learners to engage in metalinguistic reflection. As learners observe and adjust waveform patterns to match target speech, they are developing diagnostic and strategic thinking skills. These are the same skills needed in academic contexts for revision, peer feedback, and reflective practice. This supports Arifah's (2014) claim that learning with technology stimulates higher-order cognitive functions by requiring learners to make sense of abstract input.

Furthermore, augmented reality (AR) and AI-based pronunciation tutors present challenges that simulate real-world communication problems. For instance, learners may need to adjust their speech to be intelligible to a digital interlocutor in real-time, thus requiring adaptive thinking. These tasks replicate problem-solving situations and encourage

learners to use their linguistic resources strategically, an important skill in spontaneous spoken interaction.

Learners engaged in digital pronunciation training are also involved in goal setting, performance evaluation, and the adjustment of learning strategies, all metacognitive skills that fall under HOTS. As highlighted by Tsai (2019), learners using collaborative pronunciation tools demonstrated the ability to analyze peer feedback, evaluate output quality, and synthesize improvements in pronunciation.

Ultimately, integrating HOTS into pronunciation instruction through technology creates a richer, more intellectually engaging learning experience. Rather than passively repeating sounds, learners actively construct knowledge about language, make informed decisions, and develop critical awareness of their communicative competence. This not only deepens their understanding of pronunciation but also prepares them for complex communication tasks in academic and professional settings.

1.2.7. CAPT for young learners

The use of computers for educational purposes with children has increased rapidly, and this includes language learning and pronunciation skills. For instance, the British Council's Learn English Kids website features a dedicated pronunciation activities section to support young learners. Similarly, CAPT programs such as English for Kids (Krajka, 2001) cater to children and young adults, often incorporating task-based activities like speech-based roleplays and interactive games. Additionally, apps like Articulation Station enhance learning through voice recording and auditory feedback features, enabling learners to store and track progress, thus offering a comprehensive approach to pronunciation practice.

CAPT systems incorporating ASR technology, such as Ya-Ya Language Box (Chou, 2005) and Tell Me More Kids, have gained interest for their interactive, task-based activities and the ability to provide automated feedback at the phoneme, word, or sentence level. However, the accuracy of such feedback remains limited due to challenges in recognizing and assessing children's non-native speech compared to adult native speech (Gerosa & Giuliani, 2004; Hacker et al., 2005). While early corrective feedback is crucial to prevent error fossilization, inaccurate feedback risks negatively impacting children's pronunciation, motivation, and engagement with language learning.

Some studies have shown that children do enjoy using computers for pronunciation training (Chou, 2005; Mich et al., 2006). Moreover, Neri et al. (2008) demonstrated that using a CAPT system with a basic automatic speech recognition feature enabled young Italian learners of English to achieve short-term pronunciation improvements comparable to traditional teacher-led methods. These findings underscore the potential of well-designed and appropriately utilized technology in enhancing pronunciation instruction.

1.2.8. Recent developments in technology for pronunciation learning

Technological advancements have significantly influenced pronunciation learning and teaching. Key areas of innovation include mobile apps, chatbots, embodied conversational agents like ECAs or talking heads, gaming and simulations, robot-assisted language learning (RALL), and the effectiveness of computer-assisted pronunciation training (CAPT). Below is an explanation about recent developments in technology for pronunciation learning are referring to Pennington and Rogerson-Revell (2019).

First, Mobile-assisted language learning (MALL) has become prominent due to convenience, personalization, and interactive capabilities. Apps such as Oxford English File Pronunciation and Macmillan's Sounds have shown promise in improving learners' perception and production of sounds. Fouz-González (2015) demonstrated the success of mobile apps in enhancing vowel pronunciation through structured, app-based interventions. Moreover, advanced AI-based apps like ELSA Speak provide tailored feedback, enhancing both motivation and pronunciation accuracy.

Second are Chatbots, which simulate natural conversations and provide learners with opportunities to practice speaking. For instance, the Duolingo bot enables users to interact textually while simulating conversational contexts. Fryer and Carpenter (2006) highlighted the motivational potential of chatbots, noting their value in improving learners' motivation and engagement.

Third is ECAs, such as Baldi, a 3D animated "talking head," are used for improving pronunciation and intelligibility. These systems combine speech recognition, synthesis, and visual articulation to aid learners. For example, Ouni et al. (2005) emphasize that Baldi has proven effective in helping individuals, including those with hearing impairments, by visually demonstrating articulatory movements.

Fourth, games and simulations provide engaging and immersive language learning environments. Surface and Dierdorff (2007) identify that examples include World of Warcraft, which indirectly supports pronunciation skills through real-time verbal interactions, and specialized tools like Tactical Iraqi, which uses speech recognition to improve fluency and pronunciation accuracy in Arabic.

Fifth, Robot-Assisted Language Learning (RALL). Robots like IROBI and ROBOSEM have been introduced in educational settings, particularly in countries like South Korea. Han et al. (2005) found robots more effective than computers in teaching young learners' pronunciation, highlighting their potential to enhance motivation and interactive learning experiences.

Computer-Assisted Pronunciation Training (CAPT) systems, such as Pronunciation Power and Tell Me More, utilize Automatic Speech Recognition (ASR) to provide immediate feedback. Studies by Levis (2007) underline their potential in segmental and suprasegmental

training but stress the need for better pedagogy-driven designs. CAPT systems excel in delivering visual feedback, which research shows is vital for improving learners' prosody and articulation accuracy.

Innovative uses of general technologies like smartphones, podcasts, and social media platforms contribute to pronunciation learning. Fouz-González (2015) showed how integrating podcasts with interactive tools like Google Drive enhanced learners' pronunciation perception and production. Similarly, apps like Google Translate serve as effective tools for practicing minimal pairs and sentence fluency.

Based on the explanation above, it can be concluded that technological advancements have significantly transformed pronunciation learning by offering interactive, personalized, and immersive tools such as mobile apps, chatbots, ECAs, gaming, and CAPT systems. While these innovations enhance accessibility and engagement, challenges remain in providing accurate feedback and aligning with pedagogical goals. The effectiveness of these technologies lies in their ability to combine advanced features like ASR and visual feedback with sound instructional design, paving the way for more impactful language education.

2. Method

The research method used is a systematic review, which is used to answer the research questions that guide this study. A systematic review is an empirical method to minimize bias, while identifying, selecting, and synthesizing study summaries (Moher et al., 2015). In this study, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA principles) (Liberatiet al., 2009) and the extension of the PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols (PRISMA-P) (Moher et al., 2015) were used as a protocol-driven system to document the a-priori roadmap.

Data parameters were set to cover this study. The articles used in this study included several research articles from the last 20 years, which is from 2006 to 2024. where the use of technology in pronunciation teaching is growing very rapidly. The electronic search included the Educational Resources Information Center (ERIC) and ScienceDirect databases and only peer-reviewed journals were included to ensure the quality of the research that was part of this study.

This research focuses on studies that involve the use of technology in pronunciation. The search was limited to courses that only involved learning and teaching English from beginner to proficiency levels. Boolean search also was used which focused on the main aspects of the systematic review topic about technology use in teaching pronunciation.

Table 1. Boolean search terms

Search section	Search terms
Part 1	("English pronunciation teaching" OR "teaching English pronunciation" OR "pronunciation instruction") AND ("technology-assisted pronunciation training" OR "TAPT" OR "computer-assisted pronunciation training" OR "CAPT" OR "pronunciation technology")
Part 2	("technology" OR "digital tools" OR "computer-assisted language learning" OR "CALL" OR "mobile-assisted language learning" OR "MALL") AND ("English pronunciation" OR "pronunciation teaching" OR "phonetics training" OR "accent training" OR "spoken English") AND ("teaching" OR "pedagogy" OR "instruction" OR "education")

2.1. Coding and data extraction

According to Britten et al. (2002), coding data in a systematic review differs from coding primary research because participant data and author analysis are interpreted to provide third-order constructs. Coding helps find important texts in the data. Throughout the coding process, researchers continually reflect on bias to ensure the context and meaning of the original data authors are maintained when conducting secondary research.

This review began by systematically identifying relevant articles from two major academic databases: ERIC (n = 32) and ScienceDirect (n = 105), resulting in a total of 137 records identified. From screening in ScienceDirect, 6 records were removed due to missing titles, leaving 131 records for initial screening.

During title and abstract screening of the two databases, 104 records were excluded due to irrelevance, non-journal format (e.g., conference papers, book chapters, and theses), or duplication, leaving 27 full-text articles for eligibility assessment. After a thorough evaluation, all 27 articles met the inclusion criteria and were included in the final review.

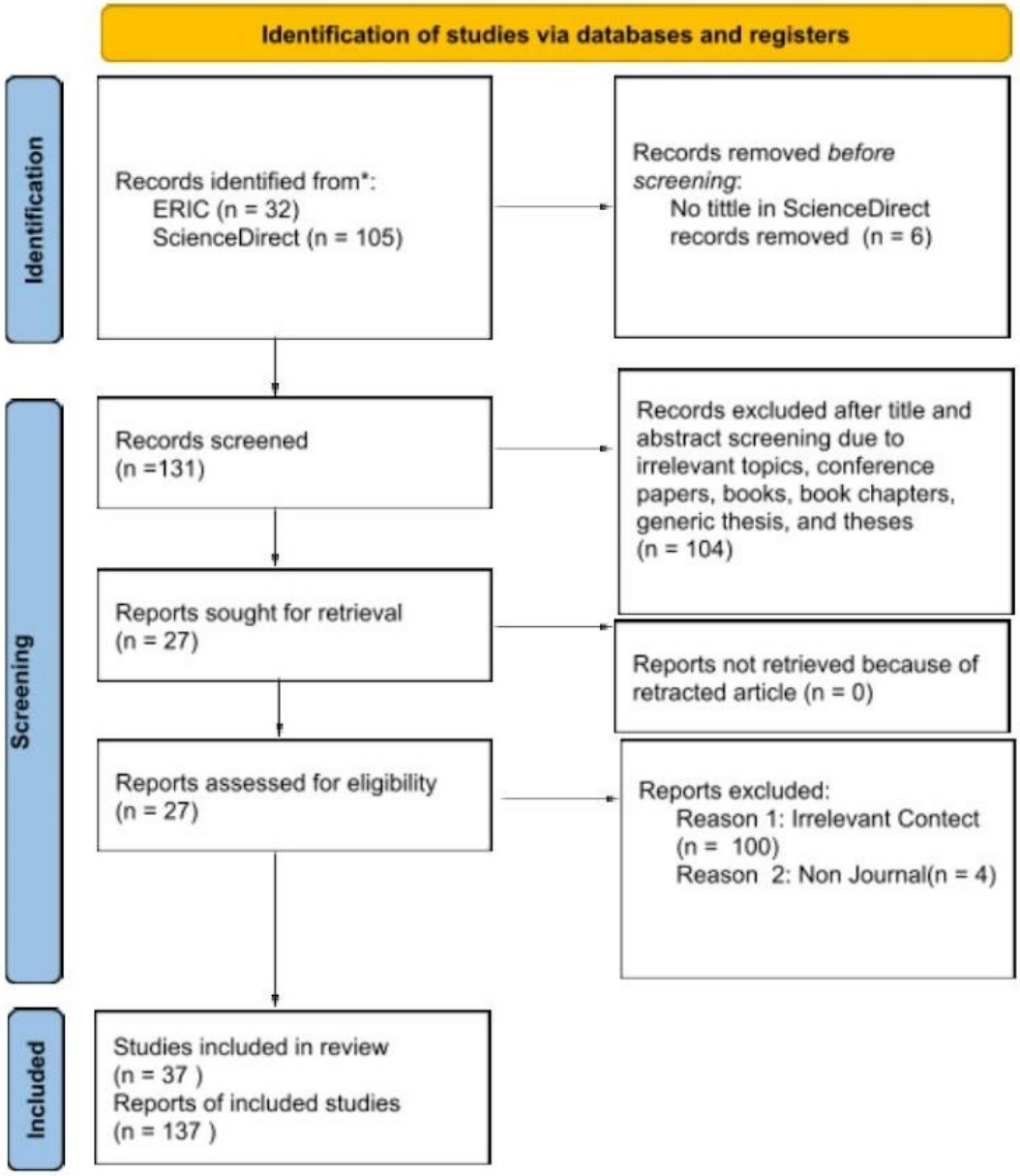

Figure 1. PRISMA Article Selection Process

Definition of Terms in the figure are according to the new PRISMA guidelines:

1. Records: Search results from bibliographic databases.
2. Reports: Full texts of papers.
3. Studies: Final decision on eligibility.

2.2. Quality assessment risk of bias

In ensuring methodological rigor of the included studies and minimizing bias, this review conducted a quality assessment using an adapted checklist based on the Critical Appraisal Skills Programme (CASP) and previous research evaluation criteria (e.g., Hsu, 2016; Amrate & Tsai, 2024). The CASP tool is a generic tool for assessing the strengths and limitations of each qualitative research methodology. The CASP tool "is designed to be used as an educational and pedagogical tool, as part of a workshop setting (CASP UK, 2025).

CASP systematically evaluates the methodological rigor of included studies, focusing on six criteria, (1) clear research objectives, (2) appropriateness of research design, (3) clear description of participants and context, (4) valid and reliable data collection tools, (5) clearly reported results, and (6) consideration of study biases or limitations. This approach follows previous applications in educational research (e.g., Long et al., 2020) and maintains the core evaluative strengths of the original CASP framework.

Each of the 27 included studies was assessed using the CASP checklist of 6 criteria for systematic reviews. Studies that met all 6 criteria were considered high quality, with clear methodology, valid tools, transparent results, and discussion of biases/limitations. Meanwhile, studies that meet 4-5 criteria are considered medium quality, where biases/limitations are less explicitly discussed. Then studies that only meet ≤3 criteria are considered low because the study does not explain important aspects such as context, design, or methods. The results of the 27 studies can be seen in the table below:

Table 2. Distribution of studies based on the CASP criteria

Quality of Studies	**Number of Studies**	**Percentage**
High	15	55.6%
Medium	11	40.7%
Low	1	3.7%
Total	27	100%

The assessment results from CAPS showed that 15 (55.6%) studies were rated as high quality, 11 (40.7%) studies were rated as moderate, and 1 (3.7%) as low based on the customized CASP checklist. This suggests that most studies had good methodology, with clear objectives, appropriate research design, and valid data collection methods. However, some studies lacked detail regarding bias or limitations.

3. Findings and Discussion

3.1. Findings

3.1.1. Types of technology

The results of the literature analysis show that the use of technology in learning and teaching pronunciation is growing rapidly from year to year where pronunciation learning has experienced a remarkable transformation with the integration of technological tools that provide innovative approaches to improving students' phonetic skills. Based on data collection the majority of research has focused on students in higher education settings, with student proficiency levels ranging from beginner to advanced.

CAPT is one of the technology systems or programs that are often used. In addition, the use of other technologies and systems used are applications and software on the current trend such as Praat software, Audacity software,Intelligent mobile application, Ichos-GSL and TheScribe, 3-D virtual talking head, FD-PSOLA (Fourier-Domain Pitch-Synchronous Overlap and Add), Golden Speaker Builder (GSB), e-learning program "Moodle (for managing learning and grading), HotPotatoes (for creating exercises), Softchalk, Online flashcard programs (e.g., Quizlet, StudyStack).

In addition, some studies use Google's speech-to-text function on a Google document, Prozed software, Pronunciation Power 2 (PP2) software, YouGlish "English Pronunciation Tutor(EPT), Pronunciation in English apps", Adaptive High-Variability Segmental Perceptual Training System/ Interface of phonetic training system, "English pronunciation training program through automatic speech recognition (ASR) computer system, Google ASR, MyET, high variability phonetic training (HVPT), virtual Language Teachers (VLT), Augmented Reality (AR), Video Conferencing Platform, Segmental Error Detection Software, PARLING (PARla INGlese, i.e.'Speak English'), and iSpraak.

Some applications and systems offer a variety of features, such as Praat and Audacity which facilitate measurable improvements in consonants, vowels and stress patterns and offer immediate phonetic feedback, allowing learners to self-review and refine their pronunciation through visual and auditory feedback. This is in line with the trend in personalized education, which increases learner engagement and effectiveness. Several areas of focus in pronunciation training with technology are also reviewed in this literature. Segmental and suprasegmental features are emphasized in all studies. Segmental aspects (e.g., consonants and vowels) are assessed in some studies, while suprasegmental features (e.g., stress and intonation) are the focus in others.

Furthermore, the interactive nature of these tools fosters deeper engagement, as students are given agency over their own practice routines. The ability to receive personalized, immediate feedback without relying entirely on instructor input empowers learners to take control of their progress. This resonates with Murray's (2005) notion that technology

cultivates learner identity and autonomy by offering natural contexts for exploration. In this regard, the tools extend beyond procedural drills to support higher-order thinking skills such as phonological analysis, error correction, and self-regulation.

In enhancing the integration of autonomous learning across devices, this review illustrates how technology can enhance learner autonomy by enabling students to receive immediate and personalized feedback, particularly through software such as Audacity, Praat, or ASR-based applications. The visual and auditory feedback mechanisms embedded in these devices encourage deeper student engagement, allowing learners to engage with themselves and to independently practice pronunciation. This aligns with the theoretical framework of technology in teaching, which emphasizes learner-centered, interactive environments.

3.1.2. Effectiveness of CAPT

Computer Assisted Pronunciation Training (CAPT) has been widely used in pronunciation instruction and training. A computer-based tool called CAPT was created to assist users with pronouncing words correctly in a second or foreign language. It frequently makes use of technologies like audio recordings, visual feedback, and speech recognition to assist students in pronouncing words more accurately and clearly. CAPT has been widely used in pronunciation instruction and training. Numerous studies have demonstrated that CAPT helps learners of all skill levels. For instance, when using CAPT beginners gain basic pronunciation skills, while advanced learners refine finer aspects such as stress and intonation.

Although CAPT is very effective, its success depends on proper integration into teaching strategies, learner motivation, and access to good technology. Current and future technological advances such as AI and virtual reality may make CAPT more interactive and efficient.

These findings support the theoretical notion that technology-mediated learning environments can enhance higher-order thinking skills through reflective practice, phonetic pattern analysis, and adaptive learning, where learning paths are dynamically adjusted based on students' abilities, learning styles, and preferences. The CAPT system's ability to respond to both familiar and novel phonetic input is described by Thomson (2011), who demonstrates how technological tools can help learners achieve more analytical and comprehensible language skills. This not only confirms the effectiveness of CAPT but also strengthens its alignment with core theoretical concepts in the use of technology in language teaching.

3.2. Discussion

The findings of this systematic review highlight the transformative impact of digital technologies on English pronunciation instruction. The reviewed studies indicate a consistent trend in the integration of tools such as Computer-Assisted Pronunciation Training (CAPT), speech analysis software (e.g., Praat and Audacity), and mobile applications, which have

contributed to enhanced learner outcomes. These technologies provide immediate feedback, support individualized practice, and accommodate learners' diverse needs across segmental and suprasegmental pronunciation features.

CAPT remains the most widely adopted tool in the reviewed literature. As demonstrated in several studies (e.g., Liu & Hung, 2016; Hsu, 2016), CAPT facilitates pronunciation training by offering repeated exposure to phonetic input, visual-auditory feedback, and self-paced learning. These tools benefit learners at various proficiency levels—from beginners acquiring basic phonemes to advanced learners refining stress and intonation. The use of visual waveforms and spectrograms, as seen in Praat-based training, also contributes significantly to learners' awareness and correction of articulation errors.

Despite the positive outcomes, the success of technology integration is influenced by contextual variables, including learners' motivation, teacher involvement, and infrastructure availability. Some studies report that learners achieved better results when technology use was supported by pedagogical guidance and regular feedback. This suggests that technological tools are most effective when embedded within structured learning environments, rather than being used independently.

The review further reveals a concentration of research within higher education contexts, with limited studies exploring the application of pronunciation technology at the primary or secondary education levels. Although systems like MyET, PARLING, and Tell Me More Kids show potential for younger learners, challenges remain in achieving accurate feedback due to ASR limitations with children's non-native speech. Studies such as Neri et al. (2008) emphasize the importance of appropriate design and feedback mechanisms for younger users.

In addition, emerging technologies such as Augmented Reality (AR), Embodied Conversational Agents (ECAs), and chatbot-based applications have expanded the possibilities for pronunciation training. Tools like Baldi and AR-based videoconferencing platforms have been shown to enhance learner engagement and improve phonetic accuracy by providing multimodal input and visual demonstration. However, further investigation is required to ensure these tools align with pedagogical objectives and can be adapted for broader classroom use.

The methodological quality of the reviewed studies also varies. Although 59.3% of studies meet high-quality criteria based on CASP, several others lack explicit reporting on research limitations or contextual factors. This variation underscores the need for greater methodological rigor in future studies to strengthen the reliability and applicability of findings across diverse educational settings.

Therefore, this review confirms that technology offers valuable support for pronunciation instruction by enhancing accessibility, interactivity, and learner autonomy. Nevertheless, sustained improvements depend on well-integrated pedagogical approaches, adequate

teacher training, and attention to contextual challenges. Further empirical research, especially in underrepresented settings such as Indonesian secondary schools, is necessary to evaluate long-term effects and inform inclusive, context-sensitive implementation of pronunciation technology tools.

4. Conclusion

With the advancement of technology in the current era of globalization, pronunciation teaching with the use of technology has become an important concern in pronunciation teaching and training. The technology used today can test soba and provide training for both segmental and suprasegmental contexts appropriately and can be adjusted. These technologies can help students improve their phonetic skills in various contexts, bridging the gap from traditional approaches. This systematic review underscores the importance of continued exploration and innovation in this transformative field.

This article, while offering a comprehensive review of technology's role in pronunciation instruction, has notable limitations. First, the analysis focuses predominantly on tools and methodologies without delving deeply into the contextual challenges of implementing these technologies in diverse educational settings, particularly those with limited resources. Additionally, the review lacks a critical assessment of the long-term effectiveness of these tools, including whether the reported improvements in pronunciation are sustained over time. The reliance on secondary data, without empirical validation in specific contexts such as Indonesian schools, further narrows the article's practical applicability and leaves critical gaps in its findings.

Future research should address these gaps by conducting longitudinal, empirical studies that evaluate the sustained impact of technological tools on learners' pronunciation skills across different educational environments. Investigations should focus on how socio-economic and cultural factors influence the adoption and effectiveness of these technologies, ensuring that solutions are inclusive and contextually relevant. Additionally, interdisciplinary efforts to design tools that align pedagogical principles with advanced technologies, such as AI and virtual reality, will be crucial in advancing pronunciation instruction to meet the diverse needs of global learners.

References

Altun, M., & Ahmad, H. K. (2021). The use of technology in English language teaching: A literature review. *International Journal of Social Sciences & Educational Studies, 8*(1), 226–232.

Amrate, M., & Tsai, P. H. (2024). Computer-assisted pronunciation training: A systematic

review. *ReCALL*, 1–21. https://doi.org/10.1017/S095834402400001X

Arifah, A. (2014). *Study on the use of technology in ELT classroom: Teachers' perspective* (Master's thesis, BRAC University). http://dspace.bracu.ac.bd/handle/10361/3449

Benson, P. (2011). *Teaching and researching autonomy* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833767

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. David McKay.

Britten, N., Campbell, R., Pope, C., Donovan, J., Morgan, M., & Pill, R. (2002). Using meta ethnography to synthesise qualitative research: A worked example. *Journal of Health Services Research & Policy, 7*(4), 209–215. https://doi.org/10.1258/135581902320432732

Cámara Arenas, E., Tejedor García, C., Tomas Vázquez, C. J., & Escudero Mancebo, D. (2024). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Computer Assisted Language Learning*. (In press)

CASP UK. (2025). *Critical Appraisal Skills Programme: Checklists and how to use them.* https://casp-uk.net/casp-tools-checklists/

Chou, F. (2005). Ya-Ya language box: A portable device for English pronunciation training with speech recognition technologies. In *Proceedings of Interspeech 2005* (pp. 169–172). http://www.isca-speech.org/archive_papers/interspeech_2005/i05_0169.pdf

Chu Liu, S., & Hung, P. Y. (2016). Teaching pronunciation with computer-assisted pronunciation instruction in a technological university. *Universal Journal of Educational Research, 4*(9), 1939–1943. https://doi.org/10.13189/ujer.2016.040901

Costley, K. C. (2014). *The positive effects of technology on teaching and student learning* [Online submission]. ERIC. https://eric.ed.gov/?id=ED554557

Cutter, M. (2015). *Using technology with English language learners in the classroom* (Master's thesis). https://digitalcommons.brockport.edu/ehd_theses/313/

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.

Eady, M. J., & Lockyer, L. (2013). Tools for learning: Technology and teaching strategies. In J. Allen (Ed.), *Learning to teach in the primary school* (pp. 71–89). Cambridge University Press.

Engwall, O., & Bälter, O. (2007). Pronunciation feedback from real and virtual language teachers. *Computer Assisted Language Learning, 20*(3), 235–262. https://doi.org/10.1080/09588220701745864

Felps, D., Ghosh, S. S., & Narayanan, S. S. (2009). A statistical model of dynamic pitch, formant, and duration features for voice conversion. *IEEE Transactions on Audio, Speech, and Language Processing, 17*(3), 529–542. https://doi.org/10.1109/TASL.2008.2007492

Fouz-González, J. (2015). *Foreign language pronunciation training with affordable and easily accessible technologies: Podcasts, smartphone apps and social networking services (Twitter)* [Doctoral dissertation, University of Murcia].

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Fryer, L., & Carpenter, R. (2006). Emerging technologies: Bots as language learning tools. *Language Learning & Technology, 10*(3), 8–14. http://llt.msu.edu/vol10num3/emerging/

Gerosa, M., & Giuliani, D. (2004). Preliminary investigations in automatic recognition of English sentences uttered by Italian children. In R. Delmonte et al. (Eds.), *Proceedings of NLP and Speech Technologies in Advanced Language Learning Systems Symposium* (pp. 9–12).

Gilakjani, A. P. (2017). A review of the literature on the integration of technology into the learning and teaching of English language skills. *International Journal of English Linguistics, 7*(5), 95–106. https://doi.org/10.5539/ijel.v7n5p95

Gilakjani, A. P., & Rahimy, R. (2019). Learners' autonomy and pronunciation learning through technology. *Journal of Language Teaching and Research, 10*(6), 1220–1226. https://doi.org/10.17507/jltr.1006.04

Gilbert, J. B. (2012). *Clear speech: Pronunciation and listening comprehension in North American English* (4th ed.). Cambridge University Press.

Gustad, A. R. (2014). The impact of technology tools on literacy motivation in elementary school English language learners: Podcasting in a 4th grade EAL class. *International Schools Journal, 34*(1), 75–84.

Hsu, L. (2016). An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning, 29*(5), 881–900. https://doi.org/10.1080/09588221.2015.1057615

Ilter, B. G. (2009). Effect of technology on motivation in EFL classrooms. *TOJDE, 10*(4), 1–10.

Krajka, J. (2001). English for Kids [Software review]. *CALICO Journal, 18*(3), 345–360. https://doi.org/10.1558/cj.v18i3.345-360

Lan, E. M. (2022). A comparative study of computer and mobile-assisted pronunciation training: The case of university students in Taiwan. *Education and Information Technologies, 27*(2), 1559–1583. https://doi.org/10.1007/s10639-021-10717-5

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202. https://doi.org/10.1017/S0267190508070098

Lin, W., & Yang, S. (2011). Exploring students' perceptions of integrating Wiki

technology and peer feedback into English writing courses. *English Teaching: Practice and Critique, 10*(2), 88–103.

Little, D. (1991). *Learner autonomy: Definitions, issues and problems*. Authentik.

Liu, S. C., & Hung, P. Y. (2016). Teaching pronunciation with computer-assisted pronunciation instruction in a technological university. *Universal Journal of Educational Research, 4*(9), 1939–1943.

Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences*, *1*(1), 31-42.

Mich, O., Neri, A., & Giuliani, D. (2006). The effectiveness of a computer-assisted pronunciation training system for young foreign language learners. In *Proceedings of CALL 2006* (pp. 135–143). http://lands.let.ru.nl/literature/neri.2006.4.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Murray, D. E. (2005). Technologies for second language literacy. *Annual Review of Applied Linguistics, 25*, 188–201. https://doi.org/10.1017/S0267190505000103

Naser, I. M., & Hamzah, M. H. (2022). Pronunciation difficulties and challenges in the field of research in Jordan. *Majallat Al-Ulum Al-Insāniyya wa al-Ijtimā'iyya, 14*(6), 140–157.

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

Nomass, B. B. (2013). The impact of using technology in teaching English as a second language. *English Language and Literature Studies, 3*(1), 111–116. https://doi.org/10.5539/ells.v3n1p111

Ouni, S., Cohen, M. M., & Massaro, D. W. (2005). Training Baldi to be multilingual: A case study for an Arabic Badr. *Speech Communication, 45*(2), 115–137. https://doi.org/10.1016/j.specom.2004.11.008

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning, 12*(5), 427–440. https://doi.org/10.1076/call.12.5.427.5693

Pennington, M. C., & Rogerson-Revell, P. (2019). Using technology for pronunciation teaching, learning, and assessment. In M. C. Pennington & P. Rogerson-Revell (Eds.), *English pronunciation teaching and research: Contemporary perspectives* (pp. 235–286). Palgrave Macmillan.

Peregoy, S. F., & Boyle, O. F. (2012). *Reading, writing, and learning in ESL: A resource book for teachers* (6th ed.). Pearson.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. M. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Computer Assisted Language Learning, 31*(7), 691–717. https://doi.org/10.1080/09588221.2018.1485704

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). Springer. https://doi.org/10.1007/978-1-4614-2018-7_7

Rodinadze, S., & Zarbazoia, K. (2012). The advantages of information technology in teaching English language. *Frontiers of Language and Teaching, 3*(5), 271–275.

Schwartz, S., & Pollishuke, M. (2013). *Creating the dynamic classroom: A handbook for teachers* (2nd ed.). Pearson Education.

Solanki, D. S., & Shyamlee, M. P. (2012). Use of technology in English language teaching and learning: An analysis. In *Proceedings of the International Conference on Language, Medias and Culture (ICLMC 2012)* (Vol. 33, pp. 150–156). IACSIT Press.

Surface, E., & Dierdorff, E. (2007). *Special operations language training software measurement of effectiveness study: Tactical Iraqi study final report*. U.S. Special Operations Command.

Thomson, R. I. (2011). Computer-assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28*(3), 744–765. https://doi.org/10.11139/cj.28.3.744-765

Tsai, P. H. (2019). Beyond self-directed computer-assisted pronunciation learning: A qualitative investigation of a collaborative approach. *Computer Assisted Language Learning, 32*(7), 713–744. https://doi.org/10.1080/09588221.2018.1535502

Warschauer, M., & Meskill, C. (2000). Technology and second language learning. In J. W. Rosenthal (Ed.), *Handbook of undergraduate second language education* (pp. 303–318). Lawrence Erlbaum Associates.

Wen, Y., Li, J., Xu, H., & Hu, H. (2023). Restructuring multimodal corrective feedback through Augmented Reality (AR)-enabled videoconferencing in L2 pronunciation teaching. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2023.2198124

Zhao, Y. (2013). Recent developments in technology and language learning: Literature review and meta-analysis. *CALICO Journal, 21*(1), 7–27. https://doi.org/10.1558/cj.v21i1.7-27