Designing AI-Enhanced Pronunciation Instruction: A Conceptual Framework Integrating Perception, Feedback, and Learner Autonomy

Awad Alshehri[1] , Abdullah Ali Altamimi[2]
[1]Imam Mohammad Ibn Saud Islamic University (IMSIU)
ORCID https://orcid.org/0000-0002-3937-2413.
[2]Department of Languages and Translation, University College of Duba, University of Tabuk
ORCID: https://orcid.org/0000-0002-7697-0535

Abstract

The topic of pronunciation has historically received limited support within second language (L2) education. One reason for the lack of support for pronunciation within L2 education is that the traditional setting of L2 classrooms is not conducive to the type of feedback that pronunciation instructors require from students. Although there are a variety of AI technologies that have been introduced to L2 education that provide benefits to language learners, the integration of these technologies into pronunciation education has been lacking. The AI-Mediated Pronunciation Learning Loop (AIM-Loop) model was created to incorporate three essential components of pronunciation education: perception, AI feedback, and learner autonomy into a framework that can be utilized to continuously facilitate the pronunciation learning of L2 students. Through reviewing studies on AI and pronunciation published between 2015 and 2025, it becomes possible to identify the mechanisms that contribute to the success of pronunciation education that utilizes AI technologies and frameworks. The results of these studies reveal that the best results occur when these AI technologies are incorporated into pronunciation education in a way that ensures that each of these three components is addressed in each pronunciation learning cycle for L2 students. By proposing the AIM-Loop, researchers can utilize AI technologies to better inform pronunciation education frameworks, and pronunciation educators can utilize this model to better incorporate artificial intelligence into their classrooms and methods of pronunciation education for their L2 students.

Keywords: artificial intelligence; automatic speech recognition; computer-assisted language learning; instructional design; learner autonomy; perception training; pronunciation pedagogy; second language acquisition.

1. Introduction

While intelligible speech is essential in communication, pronunciation is a skill that is often neglected by language teachers in second language education. Pronunciation is often neglected in favor of grammar and vocabulary, as well as due to other factors such as large class sizes or teacher expertise in pronunciation teaching (Foote et al., 2011). Furthermore, pronunciation lessons that are offered to students often do not provide the kind of feedback that is needed for pronunciation to improve; feedback that is provided to students is typically limited, inconsistent, and provided by the teacher.

Artificial intelligence platforms that use automatic speech recognition (ASR) and natural language processing (NLP) can provide feedback on pronunciation that is immediate, individualized, and outside of the teacher's capacity to provide; these technologies do not require instructors to be present for pronunciation feedback to occur, and can recognize pronunciation errors that are beyond the teachers' abilities to recognize. AI platforms for pronunciation can provide feedback on articulation, acoustics, and self-evaluations of pronunciation progress, feedback that is invisible to teachers and students alike (McCrocklin, 2016; Wei, 2023). Furthermore, studies using these artificial intelligence platforms for pronunciation have shown positive effects on the pronunciation skills of students, as well as the motivation that they exhibit towards learning those pronunciation skills. Finally, the field of computer-assisted pronunciation training (CAPT) is growing in scope and sophistication with the increasing use of these AI platforms for pronunciation improvement in the last decade.

Yet despite this promise of AI technologies for aiding the learning of pronunciation, there is a gap between the technologies and their integration within pronunciation teaching. AI technologies are mostly integrated into teaching in the form of the adoption of certain applications for tasks like automatic speech recognition (ASR), high-variability phonetic training (HVPT) platforms, and gamified pronunciation learning applications. In each of these cases, however, there is little discussion of how those technologies are to be integrated into pronunciation teaching in particular; in what ways, for instance, can pronunciation teachers best utilize the benefits that AI technologies can offer to pronunciation learners? As Burston (2015) noted of another technology-integrated language learning model, without a principled discussion of how to design pronunciation teaching around those technologies, the technology will likely have little impact upon pronunciation teaching overall.

The theoretical models that have been proposed for the development of pronunciation in L2 (models like Flege's Speech Learning Model, Levis's Intelligibility Principle, and models of the development of skills like those of Anderson, 1983) have not been developed with AI technologies in mind. Each of these models describe the ways that pronunciation can develop in general, but do not discuss the integration of AI technologies into pronunciation learning models. Thus, as a result, there is currently a gap in the development of pronunciation teaching strategies for the integration of those technologies.

To address this missing element in pronunciation learning, this paper proposes the development of the AI-Mediated Pronunciation Learning Loop (AIM-Loop). The AIM-Loop is a framework that describes the integration of three components of pronunciation learning: perception training, real-time AI feedback, and learner autonomy. Each of these components is integrated within a framework that is supported by three theories: the ICAP framework (Chi & Wylie, 2014), the Technology Acceptance Model (TAM; Davis, 1989), and the perspective of Self-Determination Theory (SDT; Deci & Ryan, 1985). Each of these theories provides a justification for the inclusion of each of the mentioned components within the learning loop, supporting the development of the AIM-Loop as a theoretically-grounded model for pronunciation learning that additionally incorporates AI technologies to support the learners.

The framework presented in this paper was developed through a synthesis of the literature on AI-assisted pronunciation, collecting 21 empirical studies that have been published between 2015 and 2025. Rather than simply performing meta-analysis on these studies to identify the efficacy of various AI pronunciation tools, this research instead investigates the various elements of these studies to develop an interpretive model for how AI can be incorporated into pronunciation instruction for L2 learners.

The paper is organized as follows. Section 2 reviews the literature on AI in language education, focusing on pronunciation and phonology. Section 3 describes the methodology used to review this literature. Section 4 presents the AIM-Loop framework and describes each of its components. Section 5 discusses the implications of the framework for language teachers, curriculum designers, and developers of pronunciation AI tools. Finally, Section 6 discusses directions for future research on this topic.

2. Literature Review

In the past decade, research into the use of AI in language education has expanded at a rapid rate. However, the contributions of AI to the area of pronunciation education have been theoretically fragmented and pedagogically underspecified. Existing research on AI and pronunciation education can be categorized into four main themes that provide the justification for the development of the AIM-

Loop model: the evolution of AI in language learning, the role of AI in pronunciation and phonological development, autonomy in the use of AI learning systems, and cognitive factors that impact the effectiveness of AI in language learning. Two additional themes help to explain the limitations of existing pronunciation learning models in relation to AI, and how the AIM-Loop model aims to address those limitations. Each of these themes contributes to a field that has amassed a significant amount of research regarding the effectiveness of AI in language learning under certain conditions, but lacks the necessary theoretical foundation to explain the factors that contribute to the effectiveness of AI in language learning under different conditions and with different groups of language learners.

2.1 From Tool Adoption to Pedagogical Design: AI in Language Learning

While the literature on AI in language learning during the first decade focused on evaluating the accuracy of the systems, the recognition of speech, and the satisfaction of the language learners using those systems (Zawacki-Richter et al., 2019), the subsequent reviews of the field have focused upon the potential that AI systems have for language learning in general (Yang & Kyun, 2022; Wang et al., 2025). Nevertheless, as Shalini Roy & Gandhimathi, 2024 suggests in their review of the literature, there is a gap in the knowledge of language teachers and language learning in general in relation to the rapid adoption of AI systems into language learning (Shalini Roy & Gandhimathi, 2025). The value of AI systems for language learning is not inherent within the systems themselves, but in the relationship between those systems and the concepts of cognition and language learning that are often taught by language teachers. As such, the introduction of AI has the potential to extend the reach of the language teacher to the language students themselves through the provision of automated feedback, but only if those systems are introduced in a manner that ensures that the language learning concepts taught by the teacher continue to remain at the center of the language learning process for each student; introducing such technology without ensuring its integration with the language learning curriculum risks shifting the language learning process towards one that is not inherently related to the goals of language learning itself.

In reviewing the literature on pronunciation in particular, then, it becomes immediately clear that one of the goals of the integration of AI into language learning in general is the provision of feedback to the language learners. Such feedback is crucial for the improvement of those language learners' pronunciation skills; however, the accuracy of that feedback as it is provided to those language learners may have little relation to the pronunciation skills that are to be learned by those students.

2.2 AI in Pronunciation and Phonology: Capability Without Design

The field of pronunciation has yielded the most contributions from AI systems in the area of ASR and HVPT. ASR systems are capable of detecting errors in the production of target phonological sounds and providing feedback to the learner (Liakin et al., 2017; McCrocklin, 2016). Studies using these types of systems have reported reductions in learner anxiety, improvements in phonological awareness, and improvements in segmental accuracy (Wei, 2023). HVPT platforms expose learners to phonological sounds from a variety of different speakers and in different contexts, which helps to improve the learner's perception of those target sounds (Bradlow, 2008; Thomson, 2019).

While HVPT and ASR systems cover the two extremes of phonological awareness - the perception and production of phonological sounds - most studies evaluating these systems have assessed each component in isolation. Reviews of the existing research report that most studies have been short-term in their design, targeted to a limited number of target phonological sounds, and that the gains in phonological ability exhibited by learners tend to dissipate after the learners have exited their technological intervention (Burston, 2015; Aryanti & Santosa, 2025). Thus, while the technology is a suitable means of delivering the feedback that is necessary for learning pronunciation, the lack of an instructional sequence is the true limitation of current pronunciation AI systems. In other words, while the field has extensively studied individual AI tools to determine their effectiveness, there is a lack of research regarding the construction of an AI system that effectively incorporates the components of perception, production, and autonomy in the acquisition of phonological skills. This is the gap that the AIM-Loop aims to address.

Learner Autonomy and Self-Regulation in AI-Enhanced Environments

Autonomous and self-regulated learning have been identified as predictors of long-term development of language skills, and digital learning environments have the potential to accommodate both aspects of these types of learning (Saito & Plonsky, 2019). More specifically, AI-based pronunciation learning platforms can enable aspects of autonomous learning by providing individual students with the tools to track their own pronunciation skills over time, and to view visual representations of their developing skills (Liu et al., 2022). These types of features are associated with the motivation theories of Self-Determination Theory, which suggests that feelings of competence and autonomy lead to the development of intrinsic motivation to learn a particular skill over time.

Despite the potential for AI-based pronunciation learning platforms to enable autonomous learning of the target language by the students, such autonomous learning also has the potential to lead to detrimental effects on deep learning of the language. For instance, autonomous learning of pronunciation skills through these platforms may lead to students disregarding the critical reflection on their pronunciation skills necessary to truly learn those sounds of the language. Instead, the students may become focused upon attaining high scores on the platform, rather than deepening their knowledge of the language. Thus, though these platforms may enable autonomous learning, teachers must actively incorporate elements into the platforms that enable students to critically reflect upon the pronunciation skills that they are learning. Though there is some research on pronunciation learning platforms, there is lacking specificity in the methods for enabling such critical reflection within the platforms. The AIM-Loop addresses this necessity for incorporating aspects of autonomous learning into the platforms by including such learning as a structured and scaffolded stage within the model.

Cognitive and Pedagogical Alignment: The Case for Theory-Driven Design

A growing body of literature suggests that the effectiveness of AI in language education depends upon its alignment with established theories of second language phonology acquisition. Cognitive skill-acquisition theory (Anderson, 1983) suggests that learners often transition from knowing how to perform a skill (declarative knowledge) to performing the skill without thinking (procedural knowledge). Schmidt's (1990) Noticing Hypothesis suggests that learners must consciously recognize differences between their native language and the target language phonemes in order to learn how to produce those phonemes correctly. AI systems can implement mechanisms to facilitate procedural knowledge acquisition and noticing of language contrasts.

Despite the application of these theories in CALL, most studies have applied these theories in a tangential manner to language education and AI integration. While many CALL studies use theories related to motivation, they do not apply these theories to the design of the AI systems. The ICAP framework (Chi & Wylie, 2014) can provide structure to CALL studies by assessing the type of engagement that the AI system creates for the language learners. For instance, ICAP can be used to determine if the AI is providing experiences that require and encourage learners to progress from Passive to Active to Constructive and Interactive learning modes. Furthermore, the Technology Acceptance Model (TAM; Davis, 1989) can be used to determine why a well-designed AI system might fail to effectively implement language learning tasks if the learners feel that the system is too complex or has no relation to the language learning tasks. Furthermore, these two models can be combined into a unifying framework to simultaneously evaluate the cognitive and motivational aspects of CALL and AI implementation. The AIM-Loop model fulfills this need by integrating elements of both ICAP and TAM models within a framework that can be applied to CALL and AI systems to evaluate their effectiveness.

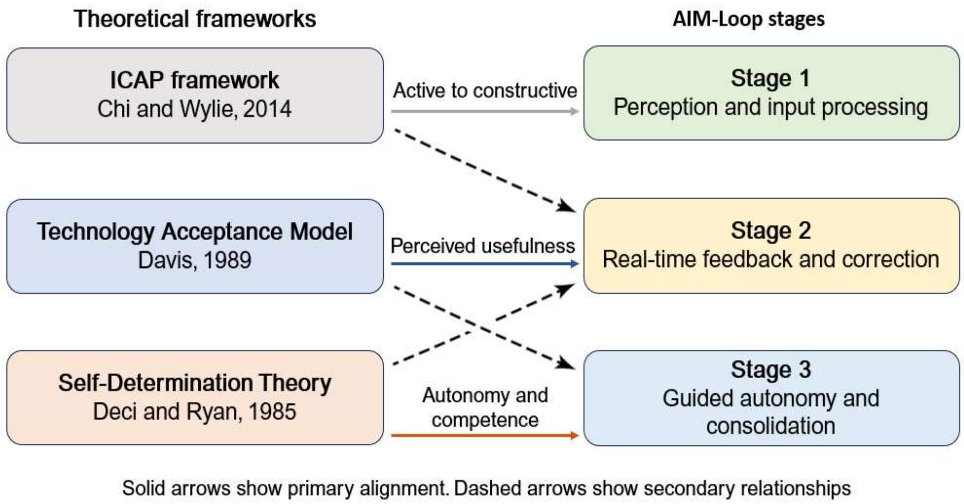

FIGURE *1*. Theoretical integration underpinning the AIM-Loop. Solid arrows indicate the primary theoretical alignment between each framework and its corresponding AIM-Loop stage. Dashed arrows indicate secondary relationships where a framework also informs adjacent stages.

Limitations of Existing Pronunciation Models in Digital Contexts

Theories of pronunciation development provide valuable cognitive and communicative foundations for understanding the learning process. However, none of these models were conceived as accounting for the affordances and constraints of AI-mediated learning environments. Flege's Speech Learning Model is detailed in describing the formation of new phonetic categories. However, the model was conceived based on naturalistic input of target phonetic sounds rather than the exposure to AI-mediated sound exposure. Levis's Intelligibility Principle and Derwing and Munro's model of communicative development emphasize the importance of listener outcomes and intelligibility in pronunciation learning, but with no description of the processes involved in the development. Models of skill acquisition, such as Anderson (1983) and Lyster's interaction framework highlight the importance of feedback in pronunciation learning, but do not describe how such feedback can be provided in an AI-mediated learning environment.

Finally, all of these models minimize the importance of motivational and affective variables in pronunciation learning. While each model describes the conditions necessary for pronunciation improvement, such as the exposure to target sounds and receiving feedback on one’s pronunciation, none of these models describe how to sustain motivation for these variables over long periods of learning, especially within an AI-mediated learning environment. Artificial intelligence technology introduces new variables and affordances to the learning process that have not been accounted for by existing models of pronunciation learning. The affordances of AI technology, such as the ability to implement multimodal feedback, track pronunciation over time, adapt to individual learners, and create emotional learning dynamics between learner and non-human feedback provider have not been recalibrated within the existing models of pronunciation development. The AIM-Loop model extends these existing models by integrating their cognitive and communicative insights with engagement and adoption theories to account for the technological affordances and new dynamics between the learner and the AI technology itself.

Persistent Design Gaps and the Rationale for the AIM-Loop

Within the themes reviewed above, three different design problems become apparent for AI-assisted pronunciation research. Firstly, the studies that exist in the field are all fragmentary in that each investigates the three components of the process separately, without considering the aspects of each as components of a designed learning process. Secondly, there is a lack of theoretical frameworks for the design of those AI-assisted pronunciation tools; few studies have considered how the technologies can relate to theories of learning, such as the ICAP, TAM, and the Noticing Hypothesis. Finally, AI-assisted pronunciation research is lacking in providing recommendations for the integration of those AI tools within pronunciation curricula for learners of different ages and backgrounds.

These three aspects of AI-assisted pronunciation research, therefore, indicate the problem that the AIM-Loop is created to address. The AIM-Loop suggests three main principles for the integration of AI tools into pronunciation teaching: that perceptual skills should precede phonetic production skills, that feedback should be provided in multiple ways and with an aim to encourage noticing of errors, and that elements of autonomy should be incorporated into encouraging learner self-regulation. Thus, the AIM-Loop framework creates an organized framework for integrating AI tools into pronunciation courses for the benefit of learners.

METHODOLOGY

Design and rationale

Given the aims of the study, employing a conceptual review methodology was both a necessary and appropriate choice of methodology. While there are numerous studies that have employed AI methodologies to investigate the use of AI in the teaching of pronunciation to second language learners, there are few studies that have attempted to provide a model for the design of pronunciation learning systems that utilize AI; such a model is lacking in the existing research. A review of the existing research on such topics is appropriate for the creation of such a model, as the aims of a conceptual review are not to calculate the effect sizes of the findings of numerous published research studies, but to recognize the concepts that have been described in those studies, to relate those concepts to existing theories within the field, and to create a model based upon those theories and concepts. Such a methodology is both reflective of the best practices in the field in regard to developing theories regarding CALL technologies (Reinders & White, 2023; Snyder, 2019), as well as reflective of the typology of literature reviews that have been proposed for research publications in the field of CAL/TAL (Paré et al., 2015).

Thus, the AI-Mediated Pronunciation Learning Loop that has been proposed and developed through this study is based upon an audit trail of the research that went into reviewing the literature on AI-enhanced pronunciation instruction for second language learners.

Corpus: Studies, Context, and Timeframe

Table 1 presents a list of the 21 empirical research studies on AI-assisted pronunciation and phonology that were included in this review. These studies were published between 2015 and 2025. This timeframe was selected for the inclusion of these studies to ensure that they referenced AI-enabled pronunciation platforms that were enabled with automatic speech recognition (ASR) technology. AI-enabled pronunciation platforms that relied upon older generations of ASR technology were specifically excluded from this review.

The studies included in this review were published in a variety of learning contexts, featured learners of diverse native language (L1) backgrounds and proficiency levels, and employed a variety of research designs. Specifically, the studies included in this review were categorized according to the nature of the AI assistance that was provided to the learners of second languages (L2). Three main categories of AI assistance to L2 pronunciation learners were identified: assistance based upon ASR technology, perception training that included high-contrast visual pronunciation training (HVPT) technologies, and autonomy/self-regulation technologies within the context of CALL (computer-assisted language learning) or computer-assisted pronunciation training (CAPT) contexts.

The decision to base this review upon these 21 research studies is explained in more detail in Section 3.4 of this study. In short, these eligibility criteria were established to ensure that the research studies that were included in the review were focused upon AI-assisted pronunciation in general, and the criteria were applied to both include and exclude certain studies. The resulting corpus of 21 studies is both dense in terms of research studies included in the review, but also because the studies that were included are all relevant to the topic of AI-assisted pronunciation for L2 learners. Each of these studies makes an important contribution to the knowledge of the mechanisms that underlie the effective integration of AI technologies in pronunciation instruction for L2 learners. Furthermore, the themes that emerged from these studies based upon these criteria were consistent with the observations from each of the individual studies. The focus on these specific criteria indicates the epistemological orientation of this study, as well as the goal of this research study in particular.

Search Strategy and Sources

A search of the Web of Science, Scopus, ERIC, and Google Scholar databases using various combinations of the following keywords was conducted: artificial intelligence AND pronunciation, ASR AND phonology instruction, perceptual training, HVPT, L2 pronunciation, computer-assisted pronunciation training, automatic speech recognition AND language learning, and adaptive feedback AND phonology. The reference lists of recent reviews in the area of interest were also mined to try to find any additional articles that might have been missed in the initial database searches. Only peer-reviewed journal articles and conference papers were included in the search. Theses, editorials, and non-peer-reviewed reports were excluded from this search.

Screening and Eligibility

Screening followed a two-pass procedure. The first pass involved reviewing the title and abstract of the studies to exclude studies that did not specifically address the following criteria: studies that did not discuss the topic of L2 pronunciation, studies that only marginally discussed the use of AI, and studies that did not include an element of education or teaching within the study. The second pass involved a thorough review of the studies to ensure that the studies met the following criteria for inclusion: the studies had to be peer-reviewed publications that were published between 2015 and 2025, they had to include some element of AI, they had to discuss the methods of the study, and their outcomes had to relate to the topic of the study of pronunciation or phonological learning. Papers that were conceptual in nature and did not contain any data, studies that published the same data as another study, and studies that did not have sufficient information regarding their methodologies were all excluded from the research. Through the screening process described above, 21 studies were included in the study for analysis.

Data Extraction

The template was applied to each of the 21 studies. For each study, the following information was recorded: the research design, the profile of the learners (who spoke what language as their L1 and at what levels of proficiency, and how many participants were involved), the instruments that were used (the ASR platforms, HVPT stimulus sets, and types of feedback provided to participants), the types of tasks that were performed by the learners (tasks like completing minimal pairs, reading aloud, having dialogues, and shadowing), the measures of the outcome of each of these tasks (segmental and prosodic accuracy and intelligibility ratings of the learners' speech), and any information regarding the motivation or affect of the participants while involved in the tasks. Additionally, information about the reliability and validity of the measurements made during each study was collected. Finally, any information regarding the visualization tools used during the studies (such as waveform and pitch visualizations of the learners' speech and dashboards that tracked their progress) was collected as well.

Analytical Procedure

Following an auditable, multi-stage procedure, the review first planned the stages of the review to define the research questions, databases, timeframes, and inclusion criteria for the studies to be reviewed, and created templates for data extraction and coding for each of the three proposed mechanisms. Following the screening and extraction stages, data extraction matrices were populated and archived. Coding sessions took place over two cycles; the first implementing codes for the three proposed mechanisms (with inductive codes additionally created from the studies to cover additional features), and the second creating broader themes from the codes created during the first cycle. Synthesis steps grouped the themes into a model that reflected each of the proposed mechanisms in the study, mapped the mechanisms to theories relating to those mechanisms, and created a model that represented the proposed mechanisms as themes, leading to the specification of the AIM-Loop. Finally, validation checks ensured the accuracy of the synthesized model and themes in comparison with other literature reviews that performed similar analyses, with a discussion of any limitations of the review in recognizing any boundary conditions for the proposed model.

Data Analysis

The analysis employed thematic synthesis combining deductive and inductive coding. Three complementary strategies strengthened analytical credibility. First, pattern matching examined whether studies that prioritized perceptual work, such as HVPT, demonstrated stronger subsequent production gains; whether visual and phonetic analytics enhanced noticing; and whether goal-tracking features correlated with persistence and self-efficacy. Second, explanatory cross-case comparison explored variation by task type, feedback granularity, and degree of teacher mediation to account for why certain mechanisms, particularly prosodic improvement, showed less consistency across studies. Third, triangulation compared outcomes across data types including acoustic measures, human ratings, and behavioral logs, across learning foci including segmental and suprasegmental features, and across contexts including laboratory, classroom, and self-access settings, to minimize bias associated with single-measure reliance.

Where numerical results were reported, effect sizes were noted descriptively but not meta-analyzed, given the heterogeneity in tasks, measures, and AI systems across the corpus. Quantitative findings were instead interpreted within thematic patterns to inform the AIM-Loop's design logic. Learning activities were mapped onto ICAP's engagement continuum from Passive through Active to Constructive and Interactive to clarify how AI tasks stimulate cognitive involvement, while TAM constructs of perceived usefulness and ease of use were applied to explain adoption and persistence, particularly in contexts where clear dashboards, interpretable feedback, and intuitive interfaces supported continued engagement.

These theoretical perspectives jointly informed the pedagogical design principles underlying the AIM-Loop: perceptual priming to reduce intrinsic cognitive load; calibrated and interpretable feedback to trigger noticing without overload; and structured self-regulation supported by goal setting, reflection prompts, and teacher mediation.

Ethical Considerations

No human participants were recruited for this study. The review draws exclusively on secondary, published sources, and accordingly IRB approval and informed consent were not required. All cited materials were used and represented in accordance with scholarly norms of accuracy and fairness. Inclusion and exclusion criteria and all analytical steps were documented transparently to enable auditability and replication.

Methodological Limitations

Three limitations to the study temper the interpretation of the review. First, because studies with unsuccessful results in AI pronunciation are less likely to be published, the proportion of studies with positive results may have been inflated. Second, because there is a lack of standardization in the measurement of speech recognition accuracy and pronunciation, it is not directly possible to compare the results of different studies. Third, the study itself is not empirical; its results are not backed by statistical analyses or experiments, but by the assertions of the study's authors. However, steps were taken to minimize this limitation by documenting all procedures, reviewing other studies on the topic, and presenting the AIM-Loop as a model that could be tested rather than a theory that should be accepted as true. These limitations are discussed in more detail in the Discussion section of the paper.

RESULTS

Overview of the Analytical Corpus

Twenty-one empirical studies were published between 2015 and 2025 and formed the basis of this review. The studies collectively involved the participation of approximately 1,600 learners from 13 different language learning contexts, the majority of which were EFL contexts located in East Asia, the Middle East, and Europe. The learners in each of these studies contributed to an average intervention duration of 5.7 weeks (SD = 2.4) and included between 24 and 160 participants per study. The studies were published in three different categories, as summarized in Table 1.

TABLE 1. Summary of Studies Included in the Conceptual Review

Category	N	Focus	Representative Measures
ASR-based feedback	9	Segmental and suprasegmental production	Pronunciation scores, intelligibility ratings
HVPT-based perception training	7	Phoneme discrimination and identification accuracy	d-prime scores, perception accuracy (%)
Mixed, integrated, or autonomy-focused designs	5	Motivation, self-efficacy, and persistence	Pre/post motivation scales, log data and use time

While the measurement of the various studies was heterogeneous, three main findings were made regarding the impact of the AIM methodology: accuracy in perceiving target sounds, pronunciation intelligibility, and the increase in the amount of self-regulated practice that occurred as a result of the methodology. The findings of these studies in both the qualitative and quantitative analyses are presented in the following sections, which lead into the three design propositions for the AIM-Loop system.

Quantitative Findings

Perceptual Gains from AI-Enhanced Input Processing

Across HVPT-based studies, the inclusion of tasks that required learners to adapt their perception of target sounds has led to mean increases in discrimination accuracy of between 18 and 27 percent relative to the accuracy of the learners' perception at baseline (M = 22.4, SD = 6.1). Furthermore, in several of the studies that have investigated these adaptations (Bradlow, 2008; Thomson, 2019), the accuracy with which participants perceived the target sounds was correlated with the accuracy with which they produced the target sounds (r = 0.42 to 0.56). These results help to suggest that the task of adapting learners' perception of target sounds is not a preparatory task for articulatory control, but instead that such a task is one of the control mechanisms for pronunciation itself. Thus, implementing technology-mediated auditory input at the beginning of pronunciation instruction, rather than as an optional component of pronunciation instruction, is a design decision that can be made based upon these results.

Production Outcomes and the Role of Feedback Design

In eight of the nine studies, participants showed significant improvement in their pronunciation after the implementation of ASR-supported practice (p less than .05). On five-point scales that measured the intelligibility of the participants' speech, the average increased 0.6 to 0.9 points. The segmental pronunciation accuracy of features like r/l sounds or th/s sounds was improved the most, while improvements in suprasegmental features like stress and rhythm were smaller but still visible. However, when waveform visualization of speech was included with the textual pronunciation feedback to the participants, studies revealed that accuracy among participants was improved an additional 8 percent on average. Thus, feedback modality impacts the success of pronunciation improvement in a way that frequency of providing feedback does not. Additionally, the limited improvements in suprasegmental features suggests a current limitation in the design of current ASRs, and provides a suggestion for the development of pronunciation courses that utilize these ASRs - they should incorporate prosodic modeling tools until the ASRs can accurately model suprasegmental features themselves.

Learner Autonomy and Motivation Indicators

Studies that included goal-setting dashboards saw significant gains in self-efficacy ($M = +0.7$ SD on five-point Likert scales) and increased voluntary practice times of 32% above their baseline. The studies also found that students' persistence with the pronunciation exercises was correlated with their perceptions of the clarity of the feedback that was provided to them in the exercises ($r = 0.48$, p less than .05). These findings validate the TAM model in that they indicate that the sense of autonomy that students have with the learning of pronunciation exercises is not automatic with simply using AI to learn pronunciation, but instead that certain features of those AI applications impact the sense of autonomy that the students have in the process of learning pronunciation. Thus, these findings have implications for the features of AI pronunciation learning platforms that are to be selected for implementation into a pronunciation curriculum.

Teacher Mediation and Instructional Design

Four quasi-experimental studies have compared the outcomes of students who used AI pronunciation tools with and without teachers interpreting the AI-generated pronunciation analyses. In each of these studies, the groups that used the AI pronunciation tools with teacher interpretation sessions achieved significantly higher intelligibility scores post-training than those who used the pronunciation tools without teacher interpretation sessions ($M = 4.3$ vs. 3.8, $t(98) = 2.64$, $p = .01$). These findings indicate that AI pronunciation tools are most effective when they are embedded within a language teaching model that utilizes teachers to interpret the AI-generated feedback to learners - rather than as self-access pronunciation training tools that do not incorporate teacher-mediated feedback.

Qualitative and Thematic Findings

Theme 1: Perceptual Priming as a Design Prerequisite

Various learners from different studies noted that the early stages of HVPT were foundational rather than supplementary, as they often stated something along the lines of "I started hearing sounds that I did not notice before." Additionally, teachers noted that learners who had completed the perception stage of HVPT were able to have smoother and more productive pronunciation sessions. Furthermore, these findings are in alignment with the theories of cognitive load, as the cognitive load of learning how to articulate a language is reduced when the learner has spent time learning to recognize phonemes. Finally, these findings help to establish the logic of the AIM-Loop in that the acquisition of perception is the first and enabling stage of HVPT rather than an additional stage to the loop.

Theme 2: Feedback as an Interactive Design Element

Learners tend to value AI feedback most when it is interpretability and actionable. Comments that reflect this include “it shows me what is wrong but also how to fix it” and “I try again until the bar turns green,” the latter of which refers to the interaction with the feedback interface. These types of interactions relate to the Constructive and Interactive levels of the ICAP model. Thus, feedback interfaces should be evaluated for how they encourage these types of interactions from the learner. Feedback that simply closes an interaction by providing a score is less valuable than feedback that opens an interaction by providing interpretable and actionable feedback regarding the target.

Theme 3: Autonomy as a Designed Outcome

Autonomy was seen in contexts where learner progress was made visible to the learner and where AI data was contextualized as data that could inform the learner about their communicative development. Learners commented on successfully using self-tracking interfaces to monitor their developing communicative skills, indicating their autonomy in such contexts: “I practiced until my curve went up.” However, however, some studies indicated that without any tasks requiring learners to reflect upon their communicative development, such self-tracking led to the misconception that rising scores indicated communicative development. Autonomy in such contexts was, therefore, strongest in contexts that utilized both the AI feedback systems as well as tasks requiring the learners to reflect upon their communicative development. Autonomy was, thus, not an inherent outcome of exposure to AI systems for developing communicative skills, but emerged only within contexts specifically designed to cultivate such autonomy. In suggesting such findings, therefore, it becomes possible to use AIM-Loop’s third stage to encourage autonomy in learners through the setting of goals, reflecting upon communicative development, and interpreting the AI data regarding communicative development at regular intervals.

Theme 4: Affective Dynamics and Feedback Tone

As pronunciation apps do not cause the same level of anxiety as teachers doing pronunciation corrections, learners can take part in pronunciation practice more willingly and frequently. However, if the app employs elements that induce frustration towards the pronunciation exercises, such as scoring systems that punish errors or gamification elements that make it so learners receive punishments for incorrect pronunciations, the affective filter will not work in the app’s advantage. Therefore, pronunciation apps should consider the affective consequences of their app designs for learners of high language anxiety.

Theme 5: The Teacher as Instructional Designer

In the various studies of teachers who used the pronunciation AI tools, the teachers began to change their roles within the classroom. The teachers began to recognize that their role was shifting from that of an instructor that corrected the students for errors to the teacher that interpreted the AI’s data and tasks for the students. This change in role is reflective of the feature of the AI tools that the teachers are supposed to utilize: the teacher’s role that the AI cannot fulfill itself is requiring the teachers to provide feedback to the students regarding the pronunciation errors that the students made, and creating activities for those students to reflect upon their errors and the AI’s suggested corrections to those errors. Thus, this theme reflects the role that teachers are to play within the AIM-Loop model - as an integral component of the model, rather than as an optional component of the process involving the AI pronunciations tools.

Integration of Quantitative and Qualitative Evidence

The convergence of quantitative outcomes and qualitative accounts across the corpus supports three design propositions that constitute the empirical foundation of the AIM-Loop.

Proposition 1: Sequenced Integration Produces Stronger Outcomes than Parallel or Unordered Deployment

Quantitative data demonstrates that learning is better with perception before production and feedback. The qualitative data explain why this is the case: the learners are ready for feedback once they have learned the auditory phonemic categories. As such, the order in which the AI tools are introduced to the learners does have an impact upon the learning process. Introducing the tools that allow for perception of the phonemic categories before the others reduces the cognitive load upon the learner and increases the quality of the learning of the phonemic categories by those learners - resulting in greater accuracy gains than if those other tools had been introduced first.

Proposition 2: Feedback Quality Determines Learning Value More Than Feedback Frequency

The statistical improvements in accuracy often correlate with the qualitative reports of the benefits of providing feedback that highlights and explains the errors of the AI models. The propositional idea that providing such feedback to the AI models will increase the language learning performance of the students has implications for the form that such a tool must take and how it may be incorporated into the curriculum for language learning programs. Overall, the value of such an AI pronunciation tool will be determined by the quality of the feedback that it provides to the students, not the amount of feedback that it provides.

Proposition 3: Autonomy Is a Designed Outcome, Not a Natural Byproduct

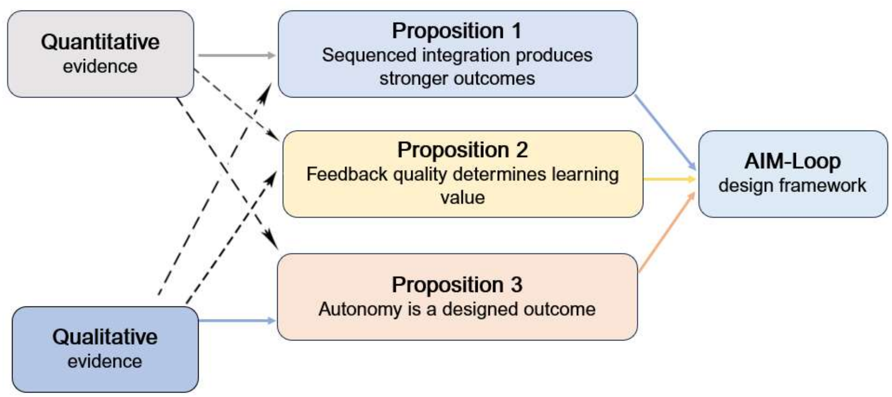

FIGURE *2.* Design propositions derived from converging evidence streams. Solid arrows indicate primary evidential support for each proposition. Dashed arrows indicate secondary contributions across propositions. All three propositions feed into the AIM-Loop design framework.

The gains in persistence and self-efficacy are made possible through the incorporation of elements that encourage and enable learners to reflect upon the outcomes of AI pronunciation tool use. Such elements include the use of data to indicate to learners of the outcomes of their use of these tools, as well as the ways in which the teachers of these tools frame the outputs of the AI tools to the learners. Overall, then, it becomes clear that elements of autonomy can be incorporated into the use of these pronunciation tool by the AI, indicating to researchers and teachers alike that such a variable can be intentionally manipulated and built by those designers and instructors of such tools.

The AIM-Loop: A Design Framework Derived from Converging Evidence

These three propositions, considered together, form the empirical and theoretical basis for the AIM-Loop. The AIM-Loop proposes that the integration of AI into pronunciation classes will occur through the implementation of three main stages. The first stage, focused on the perception of pronunciation differences, will employ AI-mediated HVPT and pronunciation discrimination tasks to build the prerequisite phonemic awareness. The second stage, focused on the provision of pronunciation feedback, will employ AI-mediated ASR to provide pronunciation feedback to the learners. Finally, the third stage consists of tasks that encourage the learners to develop autonomy in their pronunciation practice, such as implementing pronunciation reflection tasks and pronunciation goal-setting tasks, as well as periodically interpreting the pronunciation data that was collected by the AI systems. Each of these stages will feed into the next stage in a recursive manner, allowing the learners to continually develop their phonemic awareness, their ability to effectively utilize pronunciation feedback, and their self-regulatory control over their pronunciation skills.

Furthermore, the AIM-Loop is deliberately tool-agnostic; the theoretical framework for the model does not depend upon specific AI pronunciation tools, but instead upon the ability to implement such tools within the pronunciation classroom. The AIM-Loop model is, therefore, a means of transforming the literature review gap from a discussion of the individual AI tools for pronunciation into a discussion of the model that should be used to design pronunciation instruction mediated by AI tools.

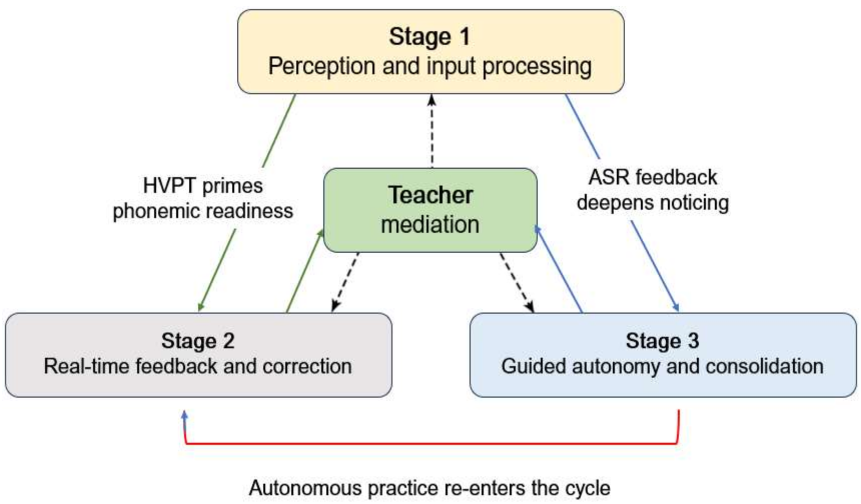

FIGURE *3*. The AI-Mediated Pronunciation Learning Loop (AIM-Loop). Dashed lines indicate teacher mediation connecting to all three stages. Solid arrows show the primary learning flow. The return arrow at the base represents the recursive re-entry of autonomous practice into the cycle.

DISCUSSION

Interpreting the Findings through a Design Lens

The findings presented in Section 4 all lead to the same conclusion: that the educational value of AI tools for pronunciation is a function of the AI tool's instructional design, not its technological features. Each of the individual AI tools has only positive impacts when incorporated into an appropriate instructional design. Thus, the findings help to answer the question of in what ways AI tools can be best utilized within pronunciation classrooms. The AIM-Loop was constructed as an answer to this question.

In the ICAP framework (Chi & Wylie, 2014), each of the three stages of the AIM-Loop can be interpreted as a stage of cognitive engagement of the learners. Each stage targets different levels of cognitive engagement so that the learners experience each of the different levels during the course of utilizing the AI tools. For instance, HVPT tasks (which can be incorporated into the Perception stage) require the learners to engage in Active and Constructive cognitive engagement. Incorporating ASR into the Pronunciation stage ensures that the learners experience Interactive cognitive engagement. Finally, the Guided Autonomy stage enables the learners to experience both Constructive and Interactive cognitive engagement after the teacher has introduced the topic and given them the tools to learn more about themselves as pronunciation learners. Thus, one measure of the success of any particular integration of AI tools into pronunciation instruction is the level of cognitive engagement that the learners experience during that incorporation of the AI tools. AI tools that do not provide opportunities for cognitive engagement beyond the Active or Passive levels are likely to have an insufficient impact on the learners' pronunciation abilities.

The Technology Acceptance Model (TAM; Davis, 1989) helps to answer the question of in what ways the learners are motivated to use these AI tools. More specifically, the model predicts that learners will be more likely to continue utilizing the AI tools if the AI tools provide clear feedback and are easy to use. The findings of the corpora indicate that the learners that continued to utilize the AI pronunciation tools after the course were the groups that found the feedback from the pronunciation tools to be interpretable and its usability to be high; those who encountered feedback that was not interpretable or whose experiences using the pronunciation tools was difficult in any way, discontinued using those tools. Thus, purchasing pronunciation AI tools based solely upon the recognition accuracy of the pronunciation tools will not ensure that the pronunciation tools will be both engaging and adopted by the students.

Convergence and Divergence with Prior Research

Overall, the findings of the present synthesis are both consistent with and extend the findings of previous research on the topic. The studies by both Liakin et al. (2017) and McCrocklin (2016) established that ASR can lead to improvements in segmental features of speech, which was again found in this current research. However, these previous studies separated the evaluation of the perception and production of those sounds and segments. The AIM-Loop research, however, indicates that the sequencing of these two processes is itself a variable that should be considered in the design of pronunciation courses that employ AI tools.

The findings of motivation to utilize AI pronunciation tools by either Vančová (2023) or Wei (2023) are also reflected in the current corpus. However, rather than finding that learners were motivated by the novelty of utilizing an AI pronunciation tool, they were motivated by the features of the tool, such as its ability to provide feedback to the learners and to observe their pronunciation errors. Thus, the motivation to use these AI pronunciation tools can be explained by those features rather than by the novelty of the technology.

In contrast to the other findings in relation to pronunciation errors, suprasegmental features tended to exhibit weaker gains. For example, several studies including that of Park (2018) found that ASR was not beneficial to the improvement of suprasegmental features like intonation and stress. One reason for this lack of benefit is that most commercially available ASR systems have a limited ability to recognize suprasegmental features of speech. Therefore, AI pronunciation courses should incorporate tools to address suprasegmental features of speech in addition to ASR systems, or until the accuracy of ASR systems in recognizing those suprasegmental features is enhanced to the point where they can be effectively utilized as a pronunciation teaching tool.

Finally, one finding that was not anticipated in the research was the affective aspect of pronunciation AI tools. The pronunciation AI tools did not always receive positive responses from the learners. For example, pronunciation AI tools that have features that may be perceived as "punitive" relative to the learners' pronunciation errors tend to create feelings of frustration and a lack of desire to utilize those pronunciation AI tools. Thus, each of the aspects of pronunciation AI tools that relate to the affective component of pronunciation learning should be considered as a variable to be considered in the design and development of pronunciation AI tools. Thus, pronunciation AI tools require the consideration and collaboration of those who consider the affective component of learning, as well as the engineers and developers of those tools.

The AIM-Loop in Relation to Existing Theoretical Models

The AIM-Loop extends rather than displaces existing theoretical accounts of L2 pronunciation development. Flege's Speech Learning Model explains how new phonetic categories form through perceptual comparison with established L1 representations, but it does not specify how adaptive digital input can accelerate or support this process. The AIM-Loop addresses this gap by positioning AI-mediated HVPT as a mechanism for delivering the varied, high-quality perceptual input that category formation requires, at a scale and consistency that naturalistic classroom exposure rarely achieves. Similarly, Levis's Intelligibility Principle foregrounds comprehensibility and listener-based outcomes as the appropriate goals of pronunciation instruction, but does not specify the instructional pathway through which learners develop the phonological control necessary to achieve those outcomes. The AIM-Loop provides that pathway by sequencing perception, feedback, and autonomy in a manner that progressively builds from phonemic awareness to accurate production to self-regulated communicative practice.

The AIM-Loop's most significant theoretical contribution, however, lies in its integration of cognitive engagement and technology adoption frameworks within a single design model. Prior pronunciation research has treated cognitive development and motivational persistence as parallel but largely independent concerns, addressed by different theoretical traditions and rarely connected at the level of instructional design. By grounding the AIM-Loop simultaneously in ICAP, TAM, and Self-Determination Theory, the framework demonstrates that cognitive engagement and motivational persistence are not independent variables to be optimized separately but interdependent design outcomes that must be addressed together. Instruction that generates high cognitive engagement but low perceived usability will not sustain the practice volume that phonological automatization requires. Instruction that sustains engagement through novelty or gamification without generating genuine cognitive demand will not produce the accuracy gains that communicative intelligibility demands. Effective design must satisfy both conditions simultaneously, and the AIM-Loop specifies how.

Pedagogical Implications

Sequenced AI Integration

The most direct implication of the AIM-Loop for pedagogy is that pronunciation classes should be organized according to a sequence of activities that does not depend upon the tools that are available to instructors, or the preferences of those instructors. Such a sequence can be implemented within any AI platform. Thus, teachers that wish to begin to teach pronunciation in AI-enhanced ways should begin with listening exercises that are similar to HVPT exercises and that target the phonemes of the learner's first language, followed by exercises in which learners pronounce phonemes with the assistance of AI and with the receiving of multimodal feedback from the AI system, and finally tasks that allow for reflection upon the pronunciation exercises and their outcomes in relation to the goals of the learners.

Teacher Mediation as Feedback Literacy

The finding that intelligibility outcomes from teacher-mediated AI sessions are significantly stronger than those from unmediated self-access with the same AI tools establishes that teacher mediation of AI pronunciation tools is a non-negotiable component of pronunciation instruction that enhances learner intelligibility. However, the role of the teacher in AI pronunciation tools is different from the role of a teacher in correcting pronunciation errors without the use of AI. Instead of the role of the teacher is to identify the errors that are detected by the AI systems and to provide feedback to the learners to correct those pronunciation errors, the teacher's role is to interpret the automated feedback that the AI system provides to the learners to develop the metacognitive and self-correction skills that are required for the learner to effectively perform such self-correction. Thus, professional development programs for teachers in institutions that implement AI pronunciation tool software should focus upon developing these feedback literacy skills in teachers to enhance the effectiveness of the pronunciation instruction that is provided to learners.

Platform Selection and Curriculum Design

When institutions consider the procurement of AI pronunciation platform tools, they should consider the criteria for design embedded within the AIM-Loop rather than considering the technical capabilities of the platform alone. Such criteria may include considerations of the interpretability and multimodality of the feedback that is provided to students, the availability of features related to tracking one's progress within the platform and setting goals, the transparency and emotional elements of the scorings and error-flagging features of the platform, and whether the platform can be configured to include teacher-mediated feedback in place of automatically-generated pronunciations. In addition to these considerations, the designers of pronunciation programs that intend to incorporate these AI tools should ensure that the tools are to be deployed within an instruction that includes pronunciation tasks, reflection tasks, and tasks that connect the AI tools to the goals that are to be accomplished by the students. Should the AI pronunciation tools be deployed without any of these surrounding instructional structures, the incorporation of such tools is likely to lead to temporary improvements in the pronunciation skills of the students alone, without any improvements in their communicative skills.

Ethical Design and Institutional Responsibility

The findings on the affective components of AI pronunciation tools indicate that these tools are not emotionally neutral. The features of these platforms can have an impact upon the emotional responses of the learners utilizing those pronunciation tools. Thus, institutions that choose to utilize these platforms are responsible for ensuring that the tools are evaluated for their potential biases, accessibility by different populations of students, and data privacy policies. Each of these components of AI pronunciation tools should be carefully considered by the institutions that implement these tools into language classrooms.

Limitations and Boundary Conditions

Three limitations to the study to which these propositions relate should be made clear. First, because most of the studies that are published regarding the use of AI in pronunciation tend to have positive outcomes, the framework's design propositions are based on a corpus of studies that is biased towards these positive outcomes. Thus, the framework has not yet been tested under scenarios in which pronunciation outcomes are less favorable, indicating the need for future studies to test these aspects of the framework. Second, the current technological limitations of ASR systems in relation to features like intonation and feedback regarding non-native accents limits the capabilities of the framework to provide effective feedback with current ASR systems. As ASR systems develop in these areas, the recommendations that are made within the framework may need to be updated to reflect these improvements in ASR systems. Third, the framework was derived from studies that were performed in EFL classrooms in East Asia, the Middle East, and Europe. Thus, the framework has not yet been tested in other parts of the world, with other native language populations, or in other classroom contexts. Future studies would be needed to determine the applicability of the AIM-Loop in other contexts than those represented in the current studies.

CONCLUSION

This paper set out to investigate the following research question: why has there been a gap in the design of pronunciation courses for SLA that incorporates artificial intelligence technology into those courses? Through reviewing 21 studies that have been published between 2015 and 2025 on the topic of the use of artificial intelligence in pronunciation courses, this review aims to answer that question and establish a framework that could be used to guide the development of pronunciation courses that utilize artificial intelligence.

The findings of the review of the existing literature on artificial intelligence in pronunciation courses establish that artificial intelligence has no inherent values for pronunciation courses; rather, the value of artificial intelligence within pronunciation courses is based upon the design of those pronunciation courses. Artificial intelligence and pronunciation learning gains are only realized in specific contexts, and if those specific contexts are not created for pronunciation courses that utilize artificial intelligence as a standalone tool, few gains can be made by incorporating artificial intelligence into pronunciation learning. Conversely, pronunciation courses that are specifically designed to include artificial intelligence as a tool that allows for pronunciation learning to take place will experience gains in pronunciation, motivation for pronunciation learning, and the development of self-regulatory skills related to pronunciation learning.

The AIM-Loop framework contributes to the existing literature on pronunciation learning and artificial intelligence in the following ways: through theoretical integration of various existing models to develop a model that explains the necessary conditions for artificial intelligence to be incorporated into pronunciation learning courses; through providing a model that can be used in future research to test the hypotheses that artificial intelligence will have positive effects upon pronunciation learning outcomes; and through modeling for pronunciation teachers, developers, and institutional stakeholders the ways in which artificial intelligence can be incorporated into pronunciation learning courses to maximize its beneficial effects.

The limitations of the review of the existing literature on artificial intelligence in pronunciation learning courses are in relation to the scope of the review itself, the methodology of the research that was reviewed, and the recommendations for future research. The review of the existing literature was limited to English as a foreign language pronunciation learning courses in regions like East Asia, the Middle East, and Europe; future studies can expand upon artificial intelligence in pronunciation learning courses to include other populations of foreign language learners with different native language backgrounds. Furthermore, the current limitations of automatic speech recognition software to recognizing prosodic elements of pronunciation learning courses suggests the need for future iterations of the AIM-Loop to incorporate pronunciation learning outcomes related to prosodic elements as pronunciation learning courses begin to utilize improved automatic speech recognition software. Finally, the knowledge of artificial intelligence in pronunciation learning courses that was gained through this literature review was based upon synthesized findings from various published studies; future studies should contribute to the existing knowledge of artificial intelligence in pronunciation learning courses through the performance of longitudinal, mixed methods studies that evaluate pronunciations learning outcomes in relation to intelligibility, and through the investigation of interactions between artificial intelligence and teachers in pronunciation learning courses.

Beyond the scope of the recommendations of future research studies, additional discussions that can be developed in the future regarding artificial intelligence in pronunciation learning courses include discussions of the potential biases of automatic speech recognition software towards non-native accents, issues of access to pronunciation learning software by all learners in pronunciation learning courses, and the effects that different types of artificial intelligence feedback have upon pronunciation learning outcomes. These topics are important topics of future discussion in the context of pronunciation learning that incorporates artificial intelligence, and should be addressed in future iterations of the AIM-Loop.

By framing artificial intelligence as a technology that can facilitate pronunciation learning courses without replacing teachers or removing pronunciation learning components and skills that must be learned, the AIM-Loop model helps to suggest a shift in the incorporation of artificial intelligence into pronunciation learning courses. Furthermore, by providing answer to the question of in what ways artificial intelligence should be incorporated into pronunciation learning courses, the AIM-Loop model helps to establish a framework for future research on the topic and artificial intelligence integration into other areas of language education.

While previous research into the use of artificial intelligence in pronunciation learning courses has mainly been limited to discussions of potential benefits to pronunciation learning courses or the components of pronunciation learning courses, the AIM-Loop model proposes that the main question of research and development efforts in pronunciation learning courses that incorporate artificial intelligence is not whether artificial intelligence can provide those benefits to pronunciation learning courses, but rather how artificial intelligence should be designed to provide such benefits. Thus, the AIM-Loop model provides an answer to that question, and can act as a foundation upon which future research and development efforts in the field can be built.

Acknowledgement

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2602)

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Aryanti, R. & Santosa, M. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language (TATEFL)*, *5*(1), 102-113. https://doi.org/10.36663/tatefl.v5i1.718

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/–/l/ contrast. https://doi.org/10.1075/sibil.36.14bra

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL, 27*(1), 4–20. https://doi.org/10.1017/S0958344014000159

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.

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

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

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal, 29*(1), 1–22. https://doi.org/10.18806/tesl.v29i1.1086

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press. https://doi.org/10.1017/9781108241564

Li, L., & Zhao, X. (2022). Investigating the integration of ASR in EFL oral practice: Students' and teachers' perspectives. *Computer Assisted Language Learning, 35*(7), 1357–1376. https://doi.org/10.1080/09588221.2021.1879921

Liakin, D., Cardoso, W., & Li, P. (2017). The use of a mobile application for L2 French pronunciation: An exploratory study. *ReCALL, 29*(1), 99–118. https://doi.org/10.1017/S095834401600018X

Liu, J., Liu, X., & Yang, C. (2022). A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*, *13*, 1049139. doi:10.3389/fpsyg.2022.1049139

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. John Benjamins.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information and Management, 52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Park, J. H. (2018). *Lexical stress sensitivity and reading in English: A systematic review and two empirical explorations among Korean-English bilinguals* (Unpublished doctoral dissertation). Texas A&M University

Thomson, R. I. (2019). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Language Teaching, 52*(4), 1–20. https://doi.org/10.1075/jslp.17038.tho
Reinders, H., & White, C. (2010). The theory and practice of technology in materials development and task design. In N. Harwood (Ed.), *English language teaching materials: Theory and practice* (pp. 58–80). Cambridge University Press
Saito, K., & Plonsky, L. (2019). *Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis*. *Language Learning, 69*(3), 652–708.
https://doi.org/10.1111/lang.12345
Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129
Shalini Roy, S., & Gandhimathi, S. N. S. (2025). Self-directed learning for optimizing sustainable language learning via mobile assisted language learning: A systematic review. *Frontiers in Education*, *9*, 1463721. https://doi.org/10.3389/feduc.2024.1463721
Snyder, H. (2019). Literature reviews as a research strategy: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Vančová, H. (2023). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, *11*(3 https://doi.org/10.2478/jolace-2023-0022
Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
Wang, Y., Wang, L., & Siau, K. L. (2025). Human-centered interaction in virtual worlds: A new era of generative artificial intelligence and metaverse. *International Journal of Human-Computer Interaction, 41*(2), 1459–1501. https://doi.org/10.1080/10447318.2024.2316376
Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology, 14*, 1261955. https://doi.org/10.3389/fpsyg.2023.1261955
Yang, H., & Kyun, S. (2022). The current research trend of artificial intelligence in language learning: A review. *Australasian Journal of Educational Technology, 38*(2), 1–18. https://doi.org/10.14742/ajet.7492
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0