

Methods to Enhance English Pronunciation for Non-English Majored Students in Vietnam: A Conceptual Framework

NGUYEN THI TUYET

Fundamentals Division, The University of Danang – Vietnam-Korea University of Information and Communication Technology, Vietnam

ABSTRACT: English pronunciation plays a critical role in communicative competence in EFL contexts; however, English pronunciation remains a persistent challenge for Vietnamese non-English majored students. Although techniques such as shadowing, phonetic training, and technology-assisted learning are widely used, they often lack theoretical integration. This conceptual paper synthesizes relevant literature to propose an Integrated Pronunciation Enhancement Model (IPEM) for them, grounded in Second Language Acquisition, Phonological Awareness, Cognitive Load Theory, and Working Memory. The model combines input enhancement, guided practice, feedback, and learner-internal factors, highlighting their interaction with cognitive processes. It systematically integrates shadowing, phonetic training, and technology-assisted learning, alongside mechanisms of attention, memory, and auditory-motor coordination. By synthesizing fragmented pedagogical approaches, the framework explains how instructional practices and cognitive constraints jointly influence pronunciation learning outcomes. The study contributes a model that supports more coherent instruction and informs future empirical research on pronunciation development in higher education.

Keywords: pronunciation; Integrated Pronunciation Enhancement Model, Vietnamese non-English majored students; conceptual framework.

Date of Submission: 05-05-2026

Date of acceptance: 16-05-2026

I. INTRODUCTION

1. Background

English pronunciation is widely recognized as a key component of communicative competence in English as a Foreign Language (EFL) contexts, particularly in relation to intelligibility and effective oral communication (Levis, 2018). In Vietnam, this issue is especially evident among non-English majored students, who often have limited exposure to authentic spoken input and receive insufficient pronunciation-focused instruction. As a result, persistent pronunciation difficulties are commonly observed, largely due to substantial phonological differences between Vietnamese and English, especially at both segmental and suprasegmental levels (Saito & Thomson, 2019).

These challenges are further reinforced by traditional instructional practices, where grammar and reading are prioritized over oral communication. Consequently, many learners develop relatively strong receptive skills but continue to struggle with spoken intelligibility and fluency (Derwing & Munro, 2015).

2. Research Gap

Despite extensive research on pronunciation teaching, several limitations remain. Firstly, instructional approaches are often examined in isolation rather than within an integrated framework. This fragmentation reduces pedagogical coherence and limits the effectiveness of instruction.

Secondly, many studies focus on short-term interventions, providing limited insight into long-term development. Third, insufficient attention has been paid to the interaction between instructional methods and cognitive processes. Finally, learner-related variables such as motivation, working memory, and prior knowledge are often overlooked.

These gaps highlight the need for a comprehensive framework that integrates instructional, cognitive, and contextual factors in pronunciation learning.

3. Purpose of the Study

This study proposes an Integrated Pronunciation Enhancement Model (IPEM) to address these limitations. The model aims to provide a theoretically grounded framework that explains how instructional methods and cognitive processes interact to facilitate pronunciation development among Vietnamese non-English majored students.

II. LITERATURE REVIEW

1. Pronunciation Challenges Among Vietnamese Non-English Majored Students

Research consistently shows that Vietnamese non-English majored students face significant pronunciation difficulties. Tran (2024) identifies word stress as the most problematic feature, followed by intonation and segmental errors. Similarly, Ha and Bao (2023) report widespread issues with final consonants, vowels, and fricatives. Tu and Du (2024) further highlight weak control of suprasegmental features affecting intelligibility.

2. Underlying Causes of Pronunciation Difficulties

Research attributes Vietnamese learners' pronunciation difficulties to interacting linguistic, cognitive, and contextual factors, with first language interference playing a central role. Vietnamese is tonal and largely monosyllabic, whereas English is stress-timed with complex syllable structures, creating systematic problems in stress, intonation, and consonant clusters (Ha and Bao, 2023). Learners must therefore adapt to a new prosodic system while acquiring unfamiliar phonetic categories. Tran (2024) also distinguishes internal factors such as awareness, motivation, and affect and external ones, including instruction and feedback. Limited exposure and exam-focused teaching further worsen persistent errors.

3. Pedagogical Approaches and Solutions

Pedagogical responses to pronunciation difficulties have explored several instructional approaches. Interactive and communicative activities, such as homophone games, have been found to improve pronunciation performance while also increasing learner engagement and motivation among non-English majored students (Minh and Giang, 2022). It underscores the value of interactive instruction. Another approach focuses on authentic materials; video-based input combined with shadowing, role-play, and guided practice helps learners improve problematic sounds, especially final consonants, through exposure to natural speech. Technology, multimedia, and corrective feedback are also integrated, while Tran (2024) highlights meaning-focused, learner-centered instruction.

4. Technology Integration in Pronunciation Learning

Technology integration in pronunciation learning has received increasing attention, as digital tools such as web-based platforms and mobile applications offer individualized practice and immediate feedback (Ngo et al., 2023; Saito, 2021). Learners generally view these environments positively due to their interactive features. However, effectiveness is still limited; pronunciation gains remain modest, and persistent errors in stress, plural endings, and linking continue. These findings suggest that technology should be embedded within a broader pedagogical framework rather than used alone.

5. Student Perspectives and Learning Strategies

Learners' perspectives provide useful insights into pronunciation learning needs. Do (2024) finds that students recognize its importance for confidence and proficiency, yet many still lack effective self-directed strategies. Phuong (2022) similarly reports limited practice time and a mismatch between teaching methods and learner preferences. Overall, these findings indicate a need for learner-centered approaches that enhance engagement and offer clear guidance for sustained pronunciation development.

6. Review of Instructional Methods

A range of instructional approaches has been proposed to address pronunciation difficulties in EFL contexts, among which shadowing, phonetic training, and technology-assisted learning are particularly prominent. Each of these methods offers distinct pedagogical advantages; however, their effectiveness varies depending on how they are implemented and integrated into the learning process.

6.1. Shadowing

Shadowing, defined as the immediate repetition of spoken input, has been increasingly recognized in recent research as an effective technique for improving both pronunciation accuracy and speaking fluency in EFL contexts (Hamada, 2024; Phan et al., 2024). Empirical studies have demonstrated that consistent engagement in shadowing activities can significantly enhance learners' pronunciation clarity, fluency, and confidence in oral communication (Lee, 2024). In particular, video-based shadowing has been shown to improve learners' control of suprasegmental features such as intonation and rhythm, which are essential for intelligible speech (Phan et al., 2024).

From a cognitive perspective, shadowing requires learners to process auditory input and produce spoken output simultaneously, thereby strengthening auditory–motor coordination and facilitating the internalization of phonological patterns. Recent systematic reviews also indicate that shadowing supports the development of bottom-up processing skills, including phoneme perception and word recognition, which are foundational for accurate pronunciation (El Moussaoui, 2025).

However, the effectiveness of shadowing is not uniform across all learners. Recent studies suggest that its success depends on factors such as learners' proficiency levels, task design, and the appropriateness of input materials (Yo Hamada, 2024). For lower-proficiency learners, the real-time processing demands of shadowing

may impose a substantial cognitive burden, potentially limiting deeper processing and resulting in surface-level repetition rather than meaningful phonological acquisition.

6.2. Phonetic Training

Phonetic training, which emphasizes explicit instruction in articulatory features and phonemic distinctions, has been widely adopted in recent pronunciation pedagogy to enhance learners' phonological awareness (Li, 2024; Mora, 2010). This approach aims to provide learners with a systematic understanding of how speech sounds are physically produced and function contrastively within the target language, thereby facilitating both perception and production of difficult phonemes. Empirical evidence from recent studies indicates that advanced forms of phonetic instruction, such as high-variability phonetic training (HVPT), can significantly improve learners' accuracy in producing specific target sounds, particularly those with clear articulatory features. (Li, 2024).

However, despite these benefits, phonetic training is often associated with considerable cognitive demands due to its technical and analytical nature (Kunová & Kováčiková, 2021). Learners are required to process abstract phonetic concepts, such as place and manner of articulation, which may impose a substantial cognitive load, particularly for lower-proficiency learners. Furthermore, recent research highlights a persistent gap between explicit phonetic knowledge and its application in spontaneous speech. While learners may demonstrate improved performance in controlled tasks, the transfer of this knowledge to real-time communication remains limited (Almusharraf, 2024). This suggests that although phonetic training enhances awareness and controlled production, it does not always guarantee automatized use in authentic communicative contexts.

6.3. Technology-Assisted Learning

In recent years, technology-assisted pronunciation learning has attracted growing scholarly attention due to rapid advancements in digital tools, mobile applications, and speech-processing technologies. Computer-assisted pronunciation training (CAPT) systems, particularly those integrating automatic speech recognition (ASR), provide learners with immediate, individualized feedback and opportunities for self-paced practice, which are often unavailable in traditional classrooms. Empirical evidence suggests that ASR-based tools can produce a moderate positive effect on learners' pronunciation performance, especially when combined with explicit corrective feedback (Ngo et al., 2023).

In addition, recent experimental studies indicate that speech-to-text technologies can enhance not only pronunciation accuracy but also learners' confidence and engagement in speaking tasks (Ngo et al., 2023). Similarly, pronunciation applications incorporating speech recognition have been shown to increase learners' motivation and willingness to communicate, largely due to the immediacy of feedback and interactive learning environments (Le & Sung, 2021). These affordances contribute to more personalized learning experiences and greater learner autonomy.

However, despite these advantages, recent literature highlights that the pedagogical effectiveness of such technologies remains contingent upon their integration into structured instructional frameworks. A growing body of research emphasizes that CAPT tools are often used as supplementary resources rather than as integral components of a coherent teaching approach ((Burston, 2015; Hubbard, 2009; Chappelle, 2001). Furthermore, many technology-driven solutions fail to adequately address individual differences in learners' proficiency levels, cognitive capacities, and learning needs, which may limit their overall instructional impact.

7. Limitations of Existing Approaches

Despite their recognized pedagogical value, existing pronunciation teaching approaches exhibit several important limitations that constrain their overall effectiveness.

Firstly, there is a notable lack of integration into a coherent instructional system. Techniques such as shadowing, phonetic training, and technology-assisted learning are often applied in isolation, without a unifying framework that connects their respective strengths. This fragmentation reduces the potential for cumulative learning effects and limits the development of a systematic pathway for pronunciation improvement.

Secondly, many approaches pay insufficient attention to the cognitive processes underlying pronunciation learning. Factors such as attention, processing capacity, and memory play a crucial role in how learners perceive, encode, and reproduce phonological information. However, instructional designs frequently overlook these cognitive dimensions, resulting in practices that may be theoretically sound but cognitively inefficient for learners.

Thirdly, there is often an imbalance between input and output in pronunciation instruction. Some methods emphasize exposure to accurate models without providing adequate opportunities for active production, while others focus heavily on repetition and output without ensuring that learners have developed sufficient perceptual awareness. Effective pronunciation development requires a balanced interaction between input and output processes, supported by meaningful practice and feedback.

Finally, learner-related variables are often underrepresented in existing approaches. Factors such as motivation, working memory capacity, and prior linguistic knowledge can significantly influence learning outcomes, yet they are rarely incorporated into instructional design. As a result, many pronunciation teaching methods fail to accommodate individual differences among learners, limiting their applicability and effectiveness in diverse educational contexts.

III. THEORETICAL FOUNDATIONS

1. Second Language Acquisition

Second Language Acquisition (SLA) theory underscores the central roles of input, interaction, and output in the language learning process (Krashen, 1985; Long, 1996; Swain, 1985). Within this framework, pronunciation development is understood as a multifaceted process that requires not only sustained exposure to accurate phonological input but also opportunities for meaningful interaction and pushed output. Exposure to high-quality input facilitates learners' ability to perceive and internalize target phonological features, while interaction provides access to feedback and negotiation of meaning, both of which are essential for phonological development (Ellis, 2008).

Moreover, the role of output is particularly critical in pronunciation learning, as it encourages learners to actively test their phonological hypotheses and refine their speech through feedback mechanisms. Empirical research has shown that opportunities for guided practice and corrective feedback significantly contribute to improvements in both segmental and suprasegmental accuracy (Saito, 2012; Thomson, 2018). Taken together, these perspectives highlight that effective pronunciation development depends on a balanced integration of input, interaction, and output within instructional contexts.

2. Phonological Awareness

Phonological awareness, defined as the ability to recognize and manipulate the sound structures of a language, constitutes a fundamental component of second language phonological development (Ehri, 2005; Lee & Mackey, 2016). In the context of pronunciation learning, phonological awareness enables learners to perceive and differentiate subtle contrasts between phonemes, which is essential for forming accurate phonological representations of the target language. Research has shown that learners with higher levels of phonological awareness demonstrate greater accuracy in both perception and production of L2 sounds, particularly when dealing with phonemic contrasts that do not exist in their first language (Best & Tyler, 2007; Saito, 2012).

Furthermore, the development of phonological awareness supports learners in establishing more robust and retrievable phonological representations, thereby facilitating more accurate and fluent speech production. Without sufficient phonological awareness, learners may struggle to distinguish similar sounds, leading to persistent pronunciation errors and reduced intelligibility in spoken communication (Derwing & Munro, 2015). These findings underscore the critical role of phonological awareness as a foundational mechanism underlying successful pronunciation acquisition.

3. Cognitive Load Theory

Cognitive Load Theory posits that learning is constrained by the limited capacity of working memory, which can be easily overwhelmed when instructional demands exceed learners' processing capabilities (Sweller, 1988; Sweller et al., 2011). Within this framework, effective instructional design requires the careful management of different types of cognitive load namely intrinsic, extraneous, and germane load to optimize learning outcomes. In particular, instructional approaches should aim to reduce unnecessary (extraneous) cognitive load while promoting germane processing that supports schema construction and automation (Paas et al., 2003).

Recent research in educational psychology continues to affirm the relevance of Cognitive Load Theory in guiding instructional practices, especially in complex learning domains such as second language acquisition, where learners must simultaneously process multiple sources of information (Kalyuga, 2020). These insights suggest that instructional design should strategically sequence tasks, provide appropriate scaffolding, and align input with learners' proficiency levels in order to prevent cognitive overload and enhance meaningful learning.

4. Working Memory

Pronunciation development in EFL contexts can be better understood through the integration of key theoretical perspectives, including Second Language Acquisition (SLA), phonological awareness, Cognitive Load Theory, and working memory. From an SLA perspective, effective learning requires not only exposure to meaningful input but also opportunities for interaction and output, which enable learners to refine their phonological representations (Ellis, 2008).

Phonological awareness plays a crucial role by allowing learners to perceive and manipulate sound structures, thereby supporting accurate pronunciation (Ehri, 2005). At the same time, Cognitive Load Theory highlights the importance of managing instructional demands, as excessive cognitive load can hinder processing efficiency (Sweller, 1988).

Within this framework, working memory is central to pronunciation learning, as it supports the temporary storage and processing of phonological input during tasks such as shadowing (Baddeley, 2000). Together, these

perspectives emphasize that pronunciation development depends on the interaction between linguistic input, cognitive processing, and instructional design.

IV. PROPOSED CONCEPTUAL FRAMEWORK

1. Overview of the Model

Building on the limitations identified in existing approaches and the theoretical foundations discussed above, this paper proposes an Integrated Pronunciation Enhancement Model (IPEM) that synthesizes multiple instructional methods within a unified and theoretically grounded framework (Ellis, 2008; Levis, 2005). The model conceptualizes pronunciation development as a dynamic process involving the interaction of three core components: input factors, learning processes, and learning outcomes, with learner-related variables functioning as moderating influences (Krashen, 1985; Swain, 1985). Rather than treating instructional techniques as isolated practices, the IPEM emphasizes the systematic integration of pedagogical inputs and cognitive mechanisms to facilitate sustainable pronunciation improvement (Saito, 2012; Derwing & Munro, 2015).

This integrative perspective is further supported by research highlighting the importance of combining multiple sources of input, opportunities for output, and feedback mechanisms in promoting pronunciation development (Ellis, 2008). In addition, cognitive theories suggest that effective learning occurs when instructional design aligns with learners' cognitive processing capacities, particularly with regard to working memory and cognitive load constraints (Sweller, 1988; Baddeley, 2000). By integrating these perspectives, the IPEM provides a comprehensive framework that accounts for both linguistic and cognitive dimensions of pronunciation learning, thereby offering a more holistic approach than traditional single-method interventions.

2. Input Factors

Within the proposed framework, input factors constitute the foundational stage of pronunciation learning by providing learners with accurate and meaningful phonological models (Krashen, 1985; Ellis, 2008). These inputs include exposure to high-quality audio materials, such as native or near-native speech, which serve as benchmarks for pronunciation accuracy and facilitate the development of perceptual representations of target sounds (Best & Tyler, 2007).

In addition, phonetic instruction plays a critical role in enhancing learners' understanding of articulatory features and phonemic distinctions, thereby supporting the development of phonological awareness (Celce-Murcia et al., 2010; Derwing & Munro, 2015). Technology-mediated tools, including speech recognition applications and digital learning platforms, further enrich input by offering interactive and individualized learning experiences, as well as immediate feedback on learners' pronunciation performance (Saito & Thomson, 2018).

Teacher-guided modeling also contributes significantly by demonstrating appropriate pronunciation patterns and scaffolding learners' engagement with input, thereby facilitating noticing and intake of phonological features (Ellis, 2008). Collectively, these input sources establish the perceptual foundation necessary for effective pronunciation development and subsequent production accuracy.

3. Learning Processes

The second component of the model focuses on the learning processes through which input is transformed into improved pronunciation performance (Ellis, 2008). Central to this stage is the use of shadowing and repetition, which facilitate the integration of listening and speaking by requiring learners to process and reproduce speech in real time (Kadota, 2019; Arguelles, 2008). These activities engage key cognitive processes, including attention and encoding, enabling learners to internalize phonological patterns (Schmidt, 1990).

Feedback mechanisms, both immediate and delayed, are incorporated to support error detection and correction, thereby reinforcing accurate pronunciation (Ellis, 2009; Saito, 2012). Through repeated practice and gradual automatization, learners are able to consolidate their phonological knowledge and develop more fluent and natural speech production (Anderson, 1983).

This process-oriented perspective highlights the importance of active engagement and iterative refinement in pronunciation learning, where continuous interaction between perception, production, and feedback leads to sustained improvement in spoken performance.

4. Learner Moderators

The effectiveness of the learning processes is influenced by several learner-related factors that function as moderators within the model (Ellis, 2008). Motivation plays a crucial role in sustaining learners' engagement and persistence in pronunciation practice, particularly in contexts where opportunities for spoken interaction are limited (Dörnyei, 2001; Dörnyei & Ushioda, 2011).

Working memory capacity is another key factor, as it determines learners' ability to process and retain phonological information during cognitively demanding tasks such as shadowing (Baddeley, 2000; Gathercole & Baddeley, 2014). Additionally, prior knowledge, including learners' existing linguistic competence and familiarity with English phonology, shapes how effectively they can interpret input and integrate new information into existing mental representations (Schmidt, 1990; Ellis, 2006).

By incorporating these variables, the model acknowledges individual differences among learners and underscores the need for adaptive instructional design that responds to learners' cognitive capacities and motivational profiles.

5. Learning Outcomes

The final component of the IPEM concerns the outcomes of pronunciation instruction, which are conceptualized in terms of three interrelated dimensions: accuracy, fluency, and intelligibility (Levis, 2005; Derwing & Munro, 2015). Improved pronunciation accuracy reflects learners' ability to produce both segmental and suprasegmental features correctly, while enhanced fluency indicates greater ease, speed, and continuity in speech production (Saito, 2012). Most importantly, increased intelligibility represents the extent to which learners' speech can be understood by listeners, which is widely regarded as the primary goal of pronunciation instruction (Levis, 2005; Derwing & Munro, 2009).

These outcomes are not viewed as isolated endpoints but as the cumulative result of the interaction between input, cognitive processing, and learner variables within the proposed framework (Ellis, 2008). Overall, the Integrated Pronunciation Enhancement Model provides a comprehensive and coherent account of how pronunciation can be effectively developed through the strategic combination of instructional methods and cognitive principles, thereby addressing both linguistic accuracy and communicative effectiveness in spoken language.

6. Relationships in the Model

The proposed framework conceptualizes pronunciation development as a set of interrelated processes in which input factors exert a direct influence on learning processes, which subsequently lead to improved learning outcomes. High-quality input, including accurate phonological models and structured instructional support, provides the necessary foundation for learners to engage in cognitive processing and productive practice. These learning processes—such as attention, encoding, shadowing, and repetition—serve as the mechanisms through which input is transformed into internalized phonological knowledge and, ultimately, more accurate and fluent speech production.

Importantly, the strength and direction of these relationships are not uniform across learners but are moderated by individual differences, including motivation, working memory capacity, and prior knowledge. These learner-related factors influence how effectively input is processed and how efficiently learning processes are executed, thereby shaping the extent to which desired outcomes are achieved. For instance, learners with higher motivation may engage more consistently in practice, while those with greater working memory capacity may handle cognitively demanding tasks such as shadowing more effectively.

Within this framework, feedback plays a pivotal mediating role by linking learning processes to outcomes. Both immediate and delayed feedback enable learners to identify discrepancies between their production and target forms, facilitating error correction and reinforcing accurate pronunciation patterns. Through iterative cycles of practice and feedback, learners progressively refine their phonological representations, leading to sustained improvements in pronunciation accuracy, fluency, and intelligibility.

V. DISCUSSION

The proposed framework addresses several critical limitations identified in the existing literature by integrating multiple instructional approaches within a coherent and theoretically grounded model. Whereas previous studies have tended to examine pronunciation teaching methods in isolation, the present framework emphasizes the dynamic interaction among pedagogical techniques and underlying cognitive processes. In doing so, it advances a more comprehensive understanding of how pronunciation development occurs, highlighting the necessity of coordinating input, output, and feedback within a structured learning system. In particular, the model underscores the importance of balancing exposure to accurate phonological input with opportunities for active production, while also incorporating feedback mechanisms that facilitate continuous refinement of learners' pronunciation.

A key contribution of the framework lies in its holistic perspective, which aligns instructional design with established cognitive principles. By integrating insights from areas such as attention, memory, and processing capacity, the model offers a more realistic account of the constraints and affordances that shape pronunciation learning. This alignment enables the design of pedagogical interventions that are not only theoretically sound but also cognitively efficient and adaptable to diverse learner profiles. As a result, the framework provides a more effective and sustainable pathway for pronunciation development, particularly in EFL contexts where learners face both linguistic and cognitive challenges.

VI. IMPLICATIONS

1. Theoretical Implications

The proposed model offers several important contributions to the theoretical understanding of pronunciation instruction in EFL contexts. By integrating insights from Second Language Acquisition, cognitive

psychology, and pedagogical practice, the framework helps bridge a long-standing gap between theory and classroom application. It reconceptualizes pronunciation development as a dynamic process shaped by the interaction of input, cognitive processing, and learner-related variables, rather than as the outcome of isolated instructional techniques. In doing so, the model extends existing pronunciation teaching theories by emphasizing the role of cognitive constraints and the need for systematic integration of instructional components. Furthermore, it provides a conceptual foundation for future research on integrated pronunciation instruction, encouraging scholars to move beyond single-method studies and examine how multiple approaches can be combined to optimize learning outcomes.

2. Pedagogical Implications

From a pedagogical perspective, the framework offers practical guidance for improving pronunciation instruction in Vietnamese university contexts. First, it highlights the importance of incorporating structured shadowing activities into classroom practice, ensuring that these activities are carefully sequenced and aligned with learners' proficiency levels. Second, the model underscores the value of using technology to provide timely and individualized feedback, as well as to expand opportunities for autonomous practice beyond the classroom. Third, it suggests that instructional design should take into account cognitive constraints by reducing unnecessary complexity and organizing learning tasks in a way that minimizes cognitive overload. Finally, the framework advocates for the systematic integration of pronunciation instruction into TOEIC-oriented courses, rather than treating it as an optional or supplementary component. By embedding pronunciation practice within existing curricula, educators can better support the development of learners' communicative competence alongside their test performance.

Future studies should empirically test the proposed model through experimental and longitudinal designs. Researchers may also explore the role of individual differences and the effectiveness of specific instructional combinations.

VII. CONCLUSION

This paper has advanced a conceptual framework for enhancing English pronunciation among non-English majored students in Vietnam by systematically integrating multiple instructional approaches with established theoretical perspectives. Grounded in insights from Second Language Acquisition and cognitive theories of learning, the proposed model conceptualizes pronunciation development as a dynamic interaction among input quality, cognitive processing, feedback mechanisms, and learner-related variables. In doing so, it addresses key limitations in the existing literature, particularly the fragmentation of instructional methods and the insufficient consideration of cognitive and contextual factors.

Beyond its theoretical contribution, the framework offers practical guidance for educators seeking to design more effective pronunciation instruction in Vietnamese EFL classrooms. By emphasizing the integration of techniques such as shadowing, feedback, and technology-enhanced practice within a cognitively informed structure, the model provides a coherent pathway for improving learners' pronunciation accuracy, fluency, and intelligibility. At the same time, it establishes a foundation for future empirical research aimed at validating and refining the proposed relationships. Overall, this study contributes to the ongoing effort to reconceptualize pronunciation teaching as a central and systematically developed component of language education.

REFERENCES

- [1]. Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.
- [2]. Arguelles, A. (2008). *Shadowing: A powerful technique for language learning*. Polyglot Conference Presentation.
- [3]. Baddeley, A. D. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences*, 4(11), 417–423.
- [4]. Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13–34). John Benjamins.
- [5]. Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, 27(1), 4–20.
- [6]. Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.
- [7]. Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.
- [8]. Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490.
- [9]. Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
- [10]. Do, M. H. (2024). Vietnamese English-major students' views and practices on English pronunciation. *Dong Thap University Journal of Science*, 13(7), 35–45.
- [11]. Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge University Press.
- [12]. Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Routledge.
- [13]. Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188.
- [14]. Ellis, N. C. (2006). Selective attention and transfer phenomena in SLA. *Applied Linguistics*, 27(2), 164–194.
- [15]. Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.
- [16]. Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3–18.
- [17]. Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press.
- [18]. Hubbard, P. (2009). *Computer assisted language learning*. Routledge.

- [19]. Kadota, S. (2019). *Shadowing as a practice in second language acquisition*. Routledge.
- [20]. Kalyuga, S. (2020). Cognitive load theory: How many types of load. *Educational Psychology Review*, 32(1), 1–19.
- [21]. Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- [22]. Kunová, M., & Kováčiková, E. (2021). Phonetic instruction and cognitive load. *System*, 98, 102115.
- [23]. Le, V. C., & Sung, Y. T. (2021). Smartphone-based speech recognition effects. *Computer Assisted Language Learning*, 34(5–6), 742–768.
- [24]. Lee, J. (2024). Effects of shadowing activities. *Convergence Studies in English Language & Literature*, 9(2), 413–439.
- [25]. Lee, J. F. K., & Mackey, A. (2016). *The Routledge handbook of second language acquisition*. Routledge.
- [26]. Levis, J. M. (2005). Changing contexts in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377.
- [27]. Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge University Press.
- [28]. Li, Y. (2024). Enhancing L2 sound learning. *Frontiers in Psychology*, 15, 1378456.
- [29]. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- [30]. Mora, J. C., & Lev-Ari, S. (2010). High variability phonetic training. *Journal of Phonetics*, 38(3), 380–388.
- [31]. Ngo, H. K., Chen, H. J., & Lai, C. F. (2023). CAPT with ASR. *ReCALL*, 35(2), 210–227.
- [32]. Nguyen T. H. M. & Nguyen M. G. (2022). Using homophone games to improve pronunciation performance for non-English majors. *International Journal of Scientific and Research Publications*, 12(5), 509–515.
- [33]. Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design. *Educational Psychologist*, 38(1), 1–4.
- [34]. Phan Thi Tuyet, V., Ly Kim, L., & Nguyen Duy, K. (2024). Video-based shadowing. *rEFLECTIONS*, 31(3), 896–924.
- [35]. Phuong, T. T. H. (2022). Feedback on pronunciation: Vietnamese EFL teachers' beliefs and practice. *SAGE Open*, 12(3).
- [36]. Saito, K. (2012). Effects of instruction on L2 pronunciation. *TESOL Quarterly*, 46(4), 842–854.
- [37]. Saito, K. (2021). Effects of instruction on L2 pronunciation development. *Language Teaching Research*, 25(3), 1–20.
- [38]. Saito, K., & Thomson, R. I. (2018). Pronunciation teaching revisited. *Language Learning*, 68(1), 299–337.
- [39]. Saito, K., & Thomson, R. I. (2019). L2 pronunciation development. *Language Teaching*, 52(2), 1–18.
- [40]. Schmidt, R. (1990). The role of consciousness in SLA. *Applied Linguistics*, 11(2), 129–158.
- [41]. Swain, M. (1985). Communicative competence: Some roles of comprehensible input and output. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- [42]. Sweller, J. (1988). Cognitive load during problem solving. *Cognitive Science*, 12(2), 257–285.
- [43]. Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- [44]. Thai T. X. H., & Nguyen K. V. B. (2023). Common errors in pronunciation of non-English majored students at the University of Transport and Communications, Ho Chi Minh Campus. *South Asian Research Journal of Arts, Language and Literature*, 5(3), 64–70.
- [45]. Thomson, R. I. (2018). Measurement of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 29–45). Routledge.
- [46]. Tran, D. T. (2024). Pronunciation problems among Vietnamese learners. *Multidisciplinary Reviews*, 7(9), 2024224.
- [47]. Tu, L., & Du, T. (2024). Pronunciation problems among Vietnamese learners. *European Journal of English Language Teaching*, 9(5).