NON-ENGLISH MAJORS' PERCEPTIONS OF SIBILANT PRONUNCIATION ERRORS: A STUDY AT A SOUTHERN VIETNAMESE UNIVERSITY

Huynh Ai[1] , Nguyen Hoang Thuan[2]
[1]Ho Chi Minh University of Banking-HUB
[2]Lac Hong University

Abstract: *Due to cross-linguistic phonological differences, Vietnamese students of English face difficulties with sibilant consonants. This study, using mixed methods, explored non-English majors' perceptions of sibilant pronunciation errors at a southern Vietnamese university. A total of 159 undergraduates completed a perception questionnaire, and a sub-sample of 18 took part in semi-structured interviews and pronunciation production tasks. The findings showed a gap between students' declarative awareness of the importance of sibilants and their limited ability to self-monitor on a real-time basis in speech production. Higher-proficiency students had stronger first-language difficulty perception and interference relative to lower-proficiency peers (higher metalinguistic awareness versus lower procedural control). Rather than withdrawing from pronunciation, students generally took compensatory measures (e.g., slowed speech, avoided lexical items) in connection with felt pronunciation difficulties. These findings emphasize an argument for moving away from rule-oriented teaching of pronunciation to training that emphasizes strategy and enhances the association between phonological awareness and spontaneous production accuracy.*

Keywords*: Sibilant consonants, cross-linguistic phonological interference, metalinguistic awareness, pronunciation self-monitoring, compensatory pronunciation strategies.*

Received: 15/01/2026 *Reviewed: 23/02/2026* *Accepted for publication: 27/02/2026*

I. INTRODUCTION

In technical and professional contexts, intelligible English pronunciation is essential. However, in English-as-a-foreign-language (EFL) settings like Vietnam, pronunciation instruction often remains marginal compared to grammar-focused exam preparation (Le & Tran, 2024), reducing students' communicative efficiency. For Vietnamese learners, pronunciation difficulties are heavily influenced by cross-linguistic differences. The Vietnamese sound system lacks several English sibilants (e.g., /z, ʃ, ʒ, tʃ, dʒ/) and favors open syllables, leading to systematic substitution, devoicing, and final consonant omission (Ha, 2005; Tran, 2019; Do, 2021).

While previous studies have well-documented these segmental errors, they predominantly focus on error description rather than learners' perceptual awareness, self-monitoring, or coping strategies (Huynh & Nguyen, 2023; Sokyrska, 2023). This leaves a critical gap regarding how non-English majors perceive the communicative impact of their errors and navigate these challenges.

To address this gap, the present study investigates the English sibilant acquisition of 159 non-English majors at a southern Vietnamese technical university. Specifically, it examines: (1) the extent to which students' perceptual awareness of sibilant errors relates to their actual articulatory accuracy and self-monitoring mechanisms, and (2) the relationship between the perceived communicative impact of these errors and the compensatory strategies employed during speech. By exploring the alignment between declarative knowledge and procedural control, the findings aim to inform strategy-oriented, intelligibility-driven pronunciation instruction for ESP contexts.

II. RESEARCH CONTENT

2.1. Sibilant production and cross-linguistic phonological constraints

English sibilants (/s, z, ʃ, ʒ, tʃ, dʒ/) pose significant challenges for Vietnamese learners. The Vietnamese sound system lacks several voiced fricatives and affricates and strongly favors open syllables (Demirezen, 2016; Ha, 2005). According to Flege's (1995) Speech Learning Model, learners assimilate unfamiliar L2 sounds to existing native categories. Consequently, cross-linguistic influence leads to widespread and systematic substitution, omission, and devoicing patterns in Vietnamese learners' English pronunciation (Do, 2021; Tran, 2019).

2.2. Perceptual awareness and metacognitive regulation in L2 pronunciation

Beyond structural constraints, pronunciation is heavily mediated by metacognitive awareness, encompassing both declarative phonetic knowledge and procedural self-monitoring (Liu, 2024). However, processing limitations and cognitive load during connected speech often prevent explicit knowledge from translating into accurate real-time production (Kartushina et al.,

2023). The alignment between learners' perceived awareness of their pronunciation errors and their objectively measured articulatory performance remains underexplored, particularly in Vietnamese tertiary contexts.

2.3. Proficiency level and perceived difficulty

Language proficiency also complicates self-perception. Advanced learners often report greater sensitivity to phonological distinctions and L1 interference (Dao, 2024), indicating increased metalinguistic awareness rather than decreased articulatory control. However, empirical evidence systematically comparing perceived difficulty with actual production accuracy across proficiency groups remains scarce among non-English majors.

2.4. Previous studies and gaps

While previous studies thoroughly document L1-induced segmental errors (Le & Tran, 2024), they primarily adopt a descriptive orientation focusing on error frequency. Concurrently, modern pronunciation pedagogy prioritizes intelligibility and communicative strategies over native-like accuracy (Burns, 2017; Gordon & Darcy, 2019; Munro, 2021). Yet, how learners themselves perceive the communicative impact of their errors and employ compensatory strategies is rarely investigated empirically (Dao, 2018). Furthermore, prior research often relies on either surveys or phonetic analyses, lacking a coherent mixed-methods integration (Creswell & Plano Clark, 2017).

Therefore, three critical gaps persist: (1) the alignment between perceptual awareness and objective sibilant production; (2) the relationship between perceived communicative impact and compensatory strategies; and (3) the influence of proficiency on perceived difficulty versus actual articulatory control.

2.5. Methodology

2.5.1. Research design

This study employed a convergent mixed-methods design to examine the relationship between students' perceptual awareness, articulatory performance, and communicative strategies in English sibilant pronunciation. In line with Creswell and Plano Clark (2017), quantitative and qualitative data were collected during the same research phase and subsequently integrated to provide a comprehensive interpretation of the findings.

The quantitative strand consisted of a questionnaire administered to 159 students to measure their perceptual awareness of sibilant difficulties, perceived communicative impact, and reported coping strategies.

The qualitative and performance-based strands involved semi-structured interviews and pronunciation production tasks conducted with a purposively selected sub-sample of 18 participants. The interviews explored learners' metacognitive experiences and strategic adjustments, while the production tasks provided objective evidence of articulatory accuracy.

This design allowed for triangulation between self-reported perceptions and observed performance, directly addressing the perception–production relationship identified in the literature.

2.5.2. Participants and context

The participants were 159 non-English major undergraduates enrolled in compulsory General English courses at a southern Vietnamese technical university during the 2025–2026 academic year. Convenience sampling was used for questionnaire administration. From this group, 18 students were purposively selected for interviews and production tasks to ensure representation across proficiency levels.

Students were enrolled in technical disciplines including Navigation, Information Technology, Mechanical Engineering, Logistics, and Data Science. The majority were first- and second-year students.

English proficiency levels were determined based on institutional placement aligned with the CEFR framework, ranging from A2 to B1. For analytical purposes, participants were categorized into two groups: Level A (Elementary) and Level B (Pre-Intermediate). This grouping enabled comparison of perceived difficulty and articulatory performance across proficiency levels.

2.5.3. Instruments

Three instruments were used to address the research questions.

2.5.3.1. Questionnaire

The Pronunciation Perception Questionnaire consisted of two main sections using a five-point Likert scale.

The first section measured learners' perceptual awareness of sibilant difficulties and their perceived ability to recognize errors in their own speech. This section was designed to capture declarative awareness and self-monitoring perceptions.

The second section examined perceived communicative consequences of pronunciation errors and the compensatory strategies students reported using (e.g., slowing down speech, lexical avoidance). This section aimed to explore the relationship between perceived intelligibility impact and strategic behavior.

2.5.3.2. Semi-structured interviews

Semi-structured interviews were conducted with 18 participants to gain deeper insight into learners' experiences of pronunciation difficulty. The interviews focused on how students interpret their errors, how they attempt to monitor their speech during communication, and what strategies they adopt when facing articulatory challenges.

These qualitative data provided contextualized explanations for questionnaire responses and allowed examination of metacognitive processes underlying strategy use.

2.5.3.3. Pronunciation production task

To obtain objective performance data, participants in the sub-sample completed a pronunciation production task targeting six English sibilants: /s, z, ʃ, ʒ, tʃ, dʒ/. The task included both word-level and sentence-level production.

The word-level task measured isolated segmental accuracy, while the sentence-level task assessed performance under increased processing demands. This distinction enabled examination of whether articulatory accuracy declined in connected speech, reflecting potential limitations in real-time monitoring.

All responses were rated using an analytic scoring rubric by two independent raters. Inter-rater reliability was calculated to ensure scoring consistency.

2.5.3.4. Procedures of data collection and data analysis

Questionnaire data were coded and analyzed using SPSS (Version 26.0). Descriptive statistics, including means and standard deviations, were calculated to examine overall perceptual awareness, perceived communicative impact, and reported compensatory strategies. Mean scores for each construct were interpreted using a five-point Likert scale. For analytical clarity, higher mean values indicated stronger agreement and greater perceived awareness or impact. Internal comparisons across constructs were conducted to identify potential discrepancies between declarative awareness and self-reported monitoring ability.

To address Research Question 1, correlational analyses were conducted to examine the relationship between perceptual awareness scores and articulatory accuracy scores derived from the production task. Additionally, to address Research Question 2, associations between perceived communicative impact and reported compensatory strategies were examined using correlation analysis and cross-construct comparisons.

Next, interview data were audio-recorded, transcribed verbatim, and analyzed using thematic analysis. An inductive coding approach was applied to identify recurring themes related to metacognitive monitoring, perceived difficulty, L1 interference, and strategy use. The qualitative findings were used to explain and contextualize quantitative patterns, particularly cases where high perceived awareness did not correspond to high articulatory accuracy.

Finally, production task analysis, participants' word-level and sentence-level productions were rated independently by two trained raters using an analytic rubric. Accuracy scores were calculated for each participant and compared across proficiency levels. Performance differences between isolated word production and sentence-level production were examined to assess whether articulatory accuracy declined under increased cognitive load, thereby providing behavioral evidence for potential monitoring limitations.

2.5.3.5. Reliability of the instruments

The reliability of the instruments was examined to ensure measurement consistency.

For the questionnaire, internal consistency was assessed using Cronbach's alpha. The Perceptual Awareness Scale (Part A) yielded an alpha coefficient of .812, while the Communicative Impact Scale (Part B) produced an alpha of .785. Both values exceed the commonly accepted threshold of .70, indicating satisfactory internal consistency.

For the production task, inter-rater reliability was calculated using Cohen's Kappa coefficient. The obtained value ($\kappa = .72$) indicates substantial agreement between raters, supporting the reliability of the articulatory accuracy scores.

Overall, these reliability indices suggest that the instruments provided consistent and dependable measurements for both perceptual constructs and performance data.

2.6. Results and Discussion

2.6.1. The metacognitive gap: Declarative awareness vs. real-time monitoring

As shown in Table 4.1, descriptive analysis revealed a clear internal asymmetry within perceptual awareness.

Table 4.1 Descriptive Statistics for Perceptual Awareness (N = 159)

Construct	Mean	SD	Interpretation
Awareness of fluency impact	3.93	1.05	High
Awareness of listener confusion	3.87	1.02	High
Difficulty distinguishing sibilants	3.50	0.93	Moderate
L1 interference awareness	3.21	1.03	Moderate
Self-recognition of errors	2.69	0.91	Low

Students demonstrated high declarative knowledge of sibilants' communicative importance (P1, P5) but reported weak procedural self-monitoring ability (P3). Interview data confirmed this retrospective awareness. As Student 11 noted, "I know the difference between /s/ and /z/, but when I talk fast, I don't think about it." Additionally, objective accuracy dropped significantly from word-level to sentence-level ($p < .001$). Interviewees confirmed that under cognitive load, attention shifts to grammar and meaning, proving that monitoring weakens as processing demands increase.

2.6.2. The proficiency paradox

To further examine this gap, learners were divided into Level A (Elementary) and Level B (Pre-Intermediate) groups.

Table 4.2. Group Differences in Perceived Difficulty and L1 Interference

Variable	Group	N	Mean	SD
P2 (Difficulty)	Level A	75	3.16	0.88
	Level B	84	3.80	0.91
P4 (L1 Interference)	Level A	75	2.95	0.97
	Level B	84	3.45	1.01

Table 4.3. Independent Samples t-Test

Variable	t	df	Sig. (2-tailed)
P2 (Difficulty)	-4.135	157	.000
P4 (L1 Interference)	-3.156	157	.002

Tables 4.2 and 4.3 reveal a "Proficiency Paradox": Level B students reported significantly higher perceived difficulty and stronger L1 interference awareness than Level A students. As proficiency increases, metalinguistic sensitivity develops faster than articulatory control. Interviews supported this; while lower-level learners prioritized basic meaning ("If people understand me, it's okay" - S1), higher-level learners showed heightened critical awareness ("The more I learn English, the more mistakes I hear" - S6).

2.6.3. Objective articulatory accuracy: L1 filtering effects

A systematic gradient of difficulty emerged, showing a robust L1 transfer effect. The L1-present phoneme /s/ yielded the highest accuracy, while L1-absent voiced sibilants (/z/, /ʒ/, /dʒ/) suffered from widespread devoicing and substitution. Despite reporting L1 interference awareness, learners structurally entrenched these errors. As Student 14 observed, "I know /z/ must vibrate, but when I speak fast, it disappears," reinforcing that L1 filtering operates procedurally regardless of conceptual awareness.

2.6.4. The pragmatic shift

Descriptive analysis revealed a hierarchy of coping strategies. Learners prioritized behavioral compensation over affective withdrawal, indicating a "Pragmatic Shift" to maintain intelligibility. Qualitative data highlighted two underlying constraints: an attention-shift effect prioritizing lexical retrieval over form ("If I concentrate on pronunciation, I cannot think about vocabulary quickly" - S3), and classroom time pressure which limits segmental monitoring opportunities.

2.6.5. Decoupling of monitoring and communicative impact

Association between perceived listener confusion and unintelligibility ($p < .001$). However, no significant association existed between self-monitoring and unintelligibility. This reveals a crucial dissociation: learners recognize when pronunciation breakdowns hinder comprehension, yet lack the internal self-regulatory mechanisms to prevent them during production.

2.7. Discussion

The findings both confirm and extend prior research (Do, 2021; Ha, 2005; Tran, 2019) by demonstrating that awareness alone does not neutralize L1 filtering. While L1-absent sibilant errors align with Flege's (1995) and Best and Tyler's (2007) assimilation models, this study reveals that articulatory instability is heavily cognitively mediated. The significant accuracy drop under cognitive load supports recent metacognitive theories (Kartushina et al., 2023; Liu, 2024), proving that pronunciation breakdown is not merely a phonetic issue, but a processing limitation.

Furthermore, the identified "Proficiency Paradox" refines Dao's (2024) observations by showing that metalinguistic awareness and procedural control follow partially independent developmental trajectories. The "Pragmatic Shift" toward compensatory strategies also strongly supports intelligibility-driven pedagogy (Burns, 2017; Munro, 2021). Ultimately, the core challenge for Vietnamese EFL learners is not simply inaccurate articulation, but the underdeveloped integration of phonological knowledge, attentional resources, and real-time monitoring mechanisms.

2.8. Implications

The results from the present study of Vietnamese EFL pronunciation instruction have meaningful pedagogical and theoretical implications. First, considering this apparent difference between perceptual awareness and real-time articulatory control, pronunciation instruction should move beyond simple rule explanation and isolated drilling to training that better prepares monitoring under communicative load. In particular, teaching practices could include graduated tasks transitioning from controlled word-level production to semi-controlled practice of the sentence, and then spontaneous speech, which can allow the participants to preserve segmental stability while absorbing meaning.

In addition, the strong effect of L1 phonological filtering implies that there is a clear need for direct contrastive training of L1-absent voiced sibilants, such as /z/, /ʒ/, and /dʒ/. But awareness is not enough. Rather, pedagogical intervention must involve both perceptual discrimination learning, articulatory position guidance and multiple production in appropriate contexts for restructuring the phonetic categories and not surface-level suppression.

Third, the identification of the "Proficiency Paradox" highlights the fact that advanced students can need various kinds of support. Higher levels of mastery may encourage greater sensitivity to minor pronunciation variations, where further instruction at higher levels can be improved through practices in self-recording, slow playback of a work in progress, and guided self-assessment assignments that will reframe increased awareness into procedural retention.

Finally, the results on compensatory strategies demonstrate that students have a very communicative orientation. Instead of discouraging strategic adaptation, teachers may incorporate strategy training into pronunciation pedagogy to support students in mindful speech rate management, breakdown repair, and prioritization of intelligibility. This provides guidance on pronunciation instruction to be more aligned with intelligibility-based frameworks but also addresses segmental accuracy.

III. CONCLUSION

This study shows that the problems of Vietnamese EFL students in the use of English sibilants are not only the product of incorrect segmental English sibilant articulation, but also a misalignment between perceptual awareness and real-time articulatory control. While students appreciate that accurate pronunciation is vital for communication, and demonstrate sensitivity to possible listener confusion, their ways of monitoring listeners break down under high cognitive load, especially for L1-absent voiced sibilants. Furthermore, proficiency is linked with metalinguistic awareness rather than full procedural accuracy, highlighting a developmental distance between awareness and control. In general, there is evidence that pronunciation difficulties in this situation can be metacognitive and processing-based, indicating the importance for teaching elements of awareness-monitoring-communicative implementation to be integrated in instruction.

REFERENCES

Burns, A. (2017). Research and the teaching of speaking in the second language classroom. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning (Vol. 3, pp. 242–256)*. Routledge.

Dao, N. A. D. (2024). Learner perceptions of L2 pronunciation instruction: A comparative study. *Science & Technology Development Journal*, 27(SI), 17–28. https://doi.org/10.32508/stdj.v27iSI.4413

Ha, C. T. (2005). Common pronunciation problems of Vietnamese learners of English. *VNU Journal of Science, Foreign Languages,* 21(1), 35–46.

Le, T. T., & Tran, T. D. (2024). English pronunciation problems among Vietnamese learners of English: An empirical study. *European Journal of English Language Teaching*, 9(5), 78–91. https://doi.org/10.46827/ejel.v9i5.5664

Sokyrska, O. (2023). Pronunciation as an important aspect of ELT: Modern trends. *In Intellectual capital is the foundation of innovative development (Part 2, pp. 118–140)*. John Benjamins. https://doi.org/10.30890/2709-2313.2023-18-02-026