MSI Journal of
Education and Social Science

Volume: 2, Issue: 3 (July-September 2026)
Published by *MSI Publishers*
ISSN 3107-5940 (Online)

Received: 12-06-2026
Accepted: 21-08-2026
Published: 24-08-2026

Phonological Decoding Difficulties in EFL Listening: A Mixed Methods Study of Vietnamese University Students

Authors Details

Nguyen Minh Chau[1*]
Faculty of Foreign Languages, Nguyen Tat Thanh University, Ho Chi Minh city, Vietnam.

Corresponding Author: *Nguyen Minh Chau*

ABSTRACT: This study investigated phonological decoding difficulties experienced by Vietnamese EFL learners during listening comprehension, employing an explanatory sequential mixed methods design. A 28-item questionnaire measuring four subscales, namely Phoneme Perception, Word Segmentation, Prosodic Processing, and Comprehension Breakdown, was administered to 72 second-year English-major students at a Vietnamese university, followed by qualitative open-ended responses. Quantitative findings indicated that all four dimensions of phonological decoding difficulty were experienced at moderate-to-high levels (overall $M = 3.21$, $SD = 0.59$), with Comprehension Breakdown emerging as the most prominent difficulty ($M = 3.45$, $SD = 0.76$). The instrument demonstrated excellent internal consistency ($\alpha = .937$). Inter-subscale correlations revealed significant positive relationships among all dimensions, with the strongest association between Word Segmentation and Prosodic Processing ($r = .745$, $p < .001$). No significant differences were found by gender, self-rated proficiency, or years of English study. Qualitative analysis identified three primary themes, which included cascade breakdown, cognitive overload, and compensatory coping strategies. These findings contribute to the understanding of bottom-up processing challenges in EFL listening and offer pedagogical implications for phonological instruction in Vietnamese higher education contexts.

DOI: *10.5281/zenodo.22078495*

Keywords: *phonological decoding, EFL listening, mixed methods, Vietnamese learners, comprehension breakdown*

INTRODUCTION

Listening comprehension has long been recognized as one of the most challenging yet fundamental skills in second and foreign language acquisition (Vandergrift & Goh, 2012). Within the English as a Foreign Language (EFL) context, listening poses particular difficulties because learners must simultaneously decode acoustic signals, segment continuous speech into meaningful units, and construct meaning in real time (Field, 2008). Despite its centrality to language proficiency, listening remains, in Vandergrift's (2007) characterization, the "cinderella skill" that receives disproportionately less research and pedagogical attention compared to reading, writing, and speaking. This neglect is especially pronounced in Asian EFL contexts, where product-oriented listening assessment, which focuses on whether students obtain correct answers rather than diagnosing how they process input, continues to dominate classroom practice (Goh, 2008).

At the core of successful listening comprehension lies phonological decoding, which refers to the ability to convert acoustic input into recognizable linguistic representations (Cutler, 2012). This bottom-up processing component requires listeners to perceive individual phonemes, segment continuous speech into discrete words, interpret prosodic features such as stress and intonation, and integrate these processes rapidly enough to maintain comprehension (Field, 2008; Rost, 2016). For EFL learners, phonological decoding is particularly demanding because the sound system of the target language frequently diverges from that of the first language (L1), creating systematic perceptual difficulties rooted in cross-linguistic interference (Best & Tyler, 2007). When these bottom-up processes are not sufficiently automatized, they consume excessive working memory resources and create a cascade of comprehension failures that can undermine the entire listening experience (Goh, 2000).

Vietnamese learners of English face distinctive phonological challenges that have received limited empirical attention in the listening comprehension literature. Vietnamese is a tonal, syllable-timed language with a relatively simple syllable structure, whereas English is a stress-timed language characterized by vowel reduction, consonant clusters, and extensive connected speech phenomena including assimilation, elision, and linking (Nguyen &

 DOI: *10.5281/zenodo.22078495*

Macken, 2008). These typological differences create a complex web of perceptual obstacles that may impede the development of listening fluency. The Vietnamese vowel system, while rich in monophthongs, does not map neatly onto the English vowel space, and the absence of many English consonant clusters in Vietnamese phonotactics may create particular segmentation challenges. Research on Vietnamese EFL learners has predominantly focused on reading comprehension (Le, 2019) and speaking skills (Nguyen, 2021), while the specific phonological processing difficulties they encounter during listening remain under-explored.

Existing research on phonological decoding in L2 listening has identified several key areas of difficulty. At the segmental level, learners struggle to discriminate phonemic contrasts that do not exist in their L1 (Flege, 1995), a challenge exacerbated by differences in vowel inventories. At the suprasegmental level, differences in prosodic systems, including the contrast between syllable-timed Vietnamese and stress-timed English, create difficulties in word segmentation and the processing of reduced forms (Cutler, 2012). Furthermore, when bottom-up decoding fails, a cascade effect may ensue, where the listener's working memory becomes overloaded, comprehension breaks down, and the learner is unable to process subsequent input (Field, 2008; Goh, 2000). This cascade effect is particularly pernicious because it transforms a localized decoding failure into a global comprehension breakdown.

While previous studies have examined phonological processing difficulties across various L1 backgrounds (e.g., Goh, 2000; Graham, 2006; Hasan, 2000), there remains a notable paucity of research specifically targeting the phonological decoding challenges faced by Vietnamese EFL learners. Furthermore, much of the existing literature relies on either purely quantitative or purely qualitative approaches, which may fail to fully capture the multidimensional nature of this phenomenon. Mixed methods designs, which integrate the generalizability of quantitative measurement with the depth of qualitative exploration, are particularly well-suited to investigating such complex perceptual processes (Creswell & Plano Clark, 2018). Consequently, the present study adopts an explanatory sequential mixed methods design to address the following research questions.

The significance of this investigation lies in its potential to inform pedagogical practices within Vietnamese EFL contexts. By identifying and ranking specific dimensions of

 DOI: *10.5281/zenodo.22078495*

phonological decoding difficulty according to their severity, instructors can allocate instructional time more efficiently and design more effective, targeted interventions. Moreover, gaining an insight into the coping strategies that learners spontaneously deploy can facilitate the development of metacognitive strategy instruction. This approach builds upon students' existing compensatory behaviors while simultaneously redirecting maladaptive responses toward more productive alternatives.

This study addresses the following research questions:

RQ1: What phonological decoding difficulties do Vietnamese EFL university students experience in listening comprehension, and to what extent?

RQ2: How do students describe their lived experiences of phonological decoding difficulty, and what coping strategies do they employ when decoding breaks down?

LITERATURE REVIEW

Phonological Decoding in L2 Listening

Phonological decoding refers to the perceptual processes by which listeners extract linguistic information from the acoustic speech signal (Cutler, 2012). In first language listening, these processes are largely automatic; in second language listening, however, they frequently require controlled attention and conscious effort, thereby consuming cognitive resources that would otherwise be available for higher-order comprehension processes (Field, 2008). The construct can be decomposed into several interrelated subprocesses: phoneme perception, or the identification of individual speech sounds; word segmentation, or the parsing of continuous speech into lexical units; and prosodic processing, or the use of stress, rhythm, and intonation to guide comprehension (Rost, 2016).

Speech Perception Theory (Best & Tyler, 2007) and the Perceptual Assimilation Model (PAM-L2) provide a robust theoretical grounding for understanding why certain L2 phonemic contrasts are more difficult to master than others. According to PAM-L2, listeners perceive non-native sounds through the filter of their L1 phonological categories, often assimilating unfamiliar sounds into their closest L1 equivalents. When two L2 sounds map onto a single L1 category, discrimination becomes especially challenging, which is a pattern highly relevant to Vietnamese learners whose L1 vowel system differs

 DOI: *10.5281/zenodo.22078495*

substantially from English (Nguyen & Macken, 2008). Flege's (1995) Speech Learning Model further predicts that L2 sounds that are phonetically similar to L1 categories are harder to acquire than entirely novel sounds, because the existing L1 category often blocks the formation of a new, distinct category for the similar L2 sound.

This theoretical prediction has important implications for Vietnamese learners of English. Phonemic contrasts that map onto partially overlapping Vietnamese categories, such as the English /ɪ/–/iː/ distinction which Vietnamese learners may assimilate to a single long vowel category, are likely to pose greater discrimination challenges than contrasts involving sounds entirely absent from Vietnamese. Within the broader architecture of L2 listening, phonological decoding represents the foundational stage upon which all subsequent processing depends. Anderson's (1995) three-phase model of language comprehension, which comprises perception, parsing, and utilization, places phonological decoding squarely within the perception phase. If perception is compromised, parsing and utilization cannot proceed effectively, explaining why difficulties at the phonological level have disproportionate effects on overall comprehension.

Word Segmentation and Connected Speech

Word segmentation constitutes a critical bottleneck in L2 listening comprehension. Unlike written text, spoken language does not contain reliable word boundaries; instead, listeners must deploy segmentation strategies based on the prosodic and phonotactic cues of the target language (Cutler, 2012). English listeners exploit the Metrical Segmentation Strategy, using strong syllables as cues to word onsets, but this strategy is unavailable to learners whose L1 employs a different rhythmic system (Cutler & Butterfield, 1992). Vietnamese learners, accustomed to a syllable-timed language in which each syllable receives relatively equal duration and prominence, face particular difficulty applying stress-based segmentation in English. The rhythmic differences between the two languages may cause Vietnamese learners to attempt syllable-by-syllable parsing of English speech, a strategy that is inefficient for a stress-timed language in which unstressed syllables are compressed and blurred.

Connected speech phenomena further compound segmentation difficulties. In natural English speech, processes such as assimilation, elision, and linking blur or eliminate word boundaries that learners expect based on their knowledge of citation forms (Cauldwell,

 DOI: *10.5281/zenodo.22078495*

2013). Reduced forms such as “gonna” and “wanna” present additional challenges, as learners may not recognize these as variants of the full forms they have learned in class (Brown & Kondo-Brown, 2006). The discrepancy between classroom pronunciation models and authentic speech has been identified as a major source of frustration for EFL learners and a significant contributor to comprehension difficulty (Cauldwell, 2013).

Comprehension Breakdown and Cognitive Load

When phonological decoding processes falter, the consequences extend beyond the immediate perception failure. Field (2008) describes a “processing bottleneck” in which the listener’s cognitive resources become consumed by attempts to decode problematic segments, leaving insufficient capacity for the construction of meaning from subsequent input. This cascading effect, where the failure to decode one item precipitates the loss of multiple subsequent items, has been documented in several studies of L2 listening (Goh, 2000; Graham, 2006). Cognitive Load Theory (Sweller, 1988), as applied to L2 listening by Vandergrift and Goh (2012), posits that the limited capacity of working memory constrains the amount of information that can be processed simultaneously, making listeners particularly vulnerable to overload when bottom-up decoding demands are high.

Goh (2000) identified five categories of real-time processing problems reported by Chinese ESL learners, three of which, namely the inability to chunk streams of speech, not recognizing words one knows, and losing the thread, are directly related to phonological decoding. Graham (2006) similarly found that French EFL learners attributed their listening difficulties primarily to the speed of the input and the difficulty of identifying individual words. These findings point to a commonality across diverse L1 backgrounds in the experience of phonological decoding difficulty, while simultaneously suggesting that the specific manifestations may vary depending on the typological relationship between the L1 and L2 sound systems.

Phonological Challenges for Vietnamese EFL Learners

The phonological inventory of Vietnamese differs from English in several respects that create predictable decoding difficulties. Vietnamese has a relatively small consonant cluster inventory compared to English, meaning that Vietnamese learners may struggle with complex onsets and codas (Nguyen & Macken, 2008). The vowel systems differ

 DOI: *10.5281/zenodo.22078495*

considerably: Vietnamese has a larger number of monophthongs but fewer diphthongs than English, and the distribution of vowel categories does not align neatly across the two languages. Additionally, Vietnamese is a tonal language in which pitch variation signals lexical meaning, whereas English uses pitch primarily for intonation and discourse functions. This fundamental typological difference means that Vietnamese learners may attend to pitch in ways that are not helpful for English listening and may miss suprasegmental cues that signal stress patterns and phrase boundaries (Nguyen, 2021).

Furthermore, the syllable structure of Vietnamese is relatively restricted compared to English. Vietnamese syllables typically follow a (C)V(C) pattern with limited consonant clustering, whereas English permits complex consonant clusters in both onset and coda positions. This structural difference means that Vietnamese learners must process phonotactic patterns that have no counterpart in their L1, which may slow down or impair lexical access during listening. Despite these documented typological differences, empirical studies investigating how Vietnamese phonological transfer specifically manifests during EFL listening comprehension remain scarce. The present study seeks to fill this gap.

METHOD

Research Design

This study employed an explanatory sequential mixed methods design (Creswell & Plano Clark, 2018), in which quantitative data were collected and analyzed first, followed by qualitative data intended to explain and elaborate upon the quantitative findings. This design was selected because it allows the breadth of survey data to be enriched by the depth of learners' experiential accounts, providing a more comprehensive understanding of phonological decoding difficulties than either approach alone could offer.

Participants

A total of 72 second-year English-major students at a Vietnamese university participated in this study. Of these, 52 (72.2%) were female and 20 (27.8%) were male, reflecting the typical gender distribution in English-major programs in Vietnam. The majority of participants (n = 44, 61.1%) self-rated their listening proficiency at the pre-intermediate level, followed by intermediate (n = 17, 23.6%), elementary (n = 6, 8.3%), and upper-

 DOI: *10.5281/zenodo.22078495*

intermediate (n = 5, 6.9%). All participants were native speakers of Vietnamese with no reported hearing impairments. Convenience sampling was used, as participants were enrolled in English skills courses taught by the researcher. Ethical approval was obtained from the institution, and all participants provided informed consent.

Instrument

The quantitative component utilized a 28-item self-report questionnaire developed by the researcher, informed by the theoretical frameworks of Field (2008), Cutler (2012), and Vandergrift and Goh (2012). The instrument measured four dimensions of phonological decoding difficulty, each represented by a 7-item subscale: (a) Phoneme Perception (Items 1–7), (b) Word Segmentation (Items 8–14), (c) Prosodic Processing (Items 15–21), and (d) Comprehension Breakdown (Items 22–28). All items were rated on a five-point Likert scale (1 = *Never*, 2 = *Rarely*, 3 = *Sometimes*, 4 = *Often*, 5 = *Always*). The questionnaire was presented bilingually (English and Vietnamese) to ensure comprehension. Three open-ended questions were appended to the survey to elicit qualitative data regarding students' specific difficulties, the impact of prosodic features, and their coping strategies when decoding fails (see Appendix).

Data Collection Procedures

The questionnaire was administered during regular class time under the supervision of the researcher. Participants were briefed on the purpose of the study and assured that their responses would remain anonymous and confidential. No incentives were offered for participation. Students were given sufficient time to complete the survey without time pressure, and the researcher was available to clarify any questions about the items. The survey took approximately 10–15 minutes to complete. After completing the Likert-scale items, participants responded to the three open-ended questions, with the instruction to write as much or as little as they wished in either English or Vietnamese.

Validity Considerations

Content validity of the instrument was established through a systematic process. The initial item pool was developed based on an extensive review of the theoretical literature on L2 phonological processing (Field, 2008; Cutler, 2012; Vandergrift & Goh, 2012), ensuring

 DOI: *10.5281/zenodo.22078495*

that all major dimensions of phonological decoding difficulty were represented. The items were reviewed by two experienced TESOL faculty members for clarity, relevance, and alignment with the intended constructs. Based on their feedback, several items were revised for linguistic clarity and cultural appropriateness. The bilingual format was verified through a back-translation procedure to ensure equivalence between the English and Vietnamese versions. A pilot test with a small group of students (n = 10, not included in the main sample) confirmed the clarity and comprehensibility of all items.

Data Analysis

Quantitative data were analyzed using SPSS (Version 27). Internal consistency was assessed using Cronbach's alpha for each subscale and the overall instrument. Descriptive statistics (means and standard deviations) were computed at the subscale and item levels. Inter-subscale relationships were examined through Pearson correlation analysis. Non-parametric tests were employed to examine potential group differences: Mann-Whitney U tests for gender and Kruskal-Wallis tests for self-rated proficiency level. Pearson correlations were used to explore the relationship between years of English study and subscale scores.

Qualitative data from the open-ended responses were analyzed using thematic analysis following the six-phase procedure outlined by Braun and Clarke (2006). Responses were read repeatedly to achieve familiarization, initial codes were generated inductively from the data, codes were collated into potential themes, themes were reviewed and refined against the coded extracts and the full dataset, and the final thematic structure was defined and named. The analysis was conducted by the researcher, and intercoder reliability was enhanced through a process of independent coding of a subset of responses by a second coder, with discrepancies resolved through discussion. Responses written in Vietnamese were analyzed in the original language to preserve nuance and were translated into English only for reporting purposes.

The integration of quantitative and qualitative findings followed the sequential explanatory approach recommended by Creswell and Plano Clark (2018). Specifically, the qualitative themes were examined in light of the quantitative patterns to determine whether and how students' experiential accounts could explain the statistical results. This integration process

 DOI: *10.5281/zenodo.22078495*

involved identifying points of convergence, where qualitative data confirmed quantitative patterns, as well as points of expansion, where qualitative data revealed dimensions of the phenomenon not captured by the survey instrument.

FINDINGS

Instrument Reliability

Prior to the main analysis, internal consistency was evaluated using Cronbach's alpha. All four subscales demonstrated acceptable to excellent reliability: Comprehension Breakdown (α = .900), Prosodic Processing (α = .871), Word Segmentation (α = .852), and Phoneme Perception (α = .781). The overall 28-item instrument yielded α = .937, indicating excellent internal consistency (George & Mallery, 2003). Corrected item-total correlations ranged from .358 to .797, all exceeding the minimum threshold of .30 (Field, 2018), confirming that each item contributed meaningfully to its respective subscale.

Table 1: Internal Consistency of Subscales (N = 72)

Subscale	Items	α	r range
Phoneme Perception	7	.781	.375–.604
Word Segmentation	7	.852	.519–.682
Prosodic Processing	7	.871	.447–.757
Comprehension Breakdown	7	.900	.494–.797
Overall scale	28	.937	—

Quantitative Findings (RQ1)

Subscale-Level Findings.

Descriptive statistics revealed that phonological decoding difficulties were experienced at a moderate-to-high level across all four dimensions (Table 2). Comprehension Breakdown was the most severely reported difficulty (M = 3.45, SD = 0.76), followed by Phoneme Perception (M = 3.21, SD = 0.54), Word Segmentation (M = 3.10, SD = 0.67), and

 DOI: *10.5281/zenodo.22078495*

Prosodic Processing (M = 3.09, SD = 0.72). All means fell within the "sometimes–often" range, indicating that these difficulties were neither rare nor negligible in students' listening experience.

Table 2: Descriptive Statistics for Phonological Decoding Difficulty Subscales (N = 72)

Subscale	**M**	**SD**
Phoneme Perception	3.21	0.54
Word Segmentation	3.10	0.67
Prosodic Processing	3.09	0.72
Comprehension Breakdown	3.45	0.76

Note. Scale: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 5 = Always.

Item-Level Findings.

At the item level, the five most frequently reported difficulties were: contextual guessing as a compensatory response to decoding failure (Item 26: M = 3.65, SD = 0.97), difficulty with unfamiliar sound combinations (Item 4: M = 3.56, SD = 0.75), anxiety induced by decoding failure (Item 28: M = 3.54, SD = 0.92), inability to separate words during fast speech (Item 13: M = 3.49, SD = 0.84), and cascade word-loss effects (Item 27: M = 3.49, SD = 0.96). These top-ranked items span three of the four subscales, suggesting that the most severe individual difficulties reflect the interconnected nature of phonological decoding processes. Conversely, reduced forms such as "gonna" (Item 12: M = 2.72, SD = 1.02) and recognition of linked sounds (Item 9: M = 2.83, SD = 0.92) were rated least frequently, suggesting these features may be less consciously salient to learners (Cauldwell, 2013).

Inter-Subscale Correlations.

Pearson correlation analysis revealed that all four subscales were significantly and positively intercorrelated (all $p < .01$), indicating that the dimensions of phonological decoding difficulty co-occur rather than operate independently (Table 3). The strongest

 DOI: *10.5281/zenodo.22078495*

association was between Word Segmentation and Prosodic Processing ($r = .745$, $p < .001$), suggesting that difficulty parsing word boundaries is closely linked to the inability to process English stress and rhythm. Phoneme Perception showed the weakest correlation with Comprehension Breakdown ($r = .397$, $p = .001$), implying that sound-level confusion is a somewhat more discrete process from downstream meaning loss.

Table 3: *Inter-Subscale Pearson Correlations*

	S1	**S2**	**S3**	**S4**
S1. Phoneme Perception	—			
S2. Word Segmentation	.584**	—		
S3. Prosodic Processing	.512**	.745**	—	
S4. Comprehension Breakdown	.397**	.581**	.633**	—

Note. ** $p < .01$.

Effects of Demographic Variables.

Mann-Whitney U tests indicated no statistically significant differences between male and female students on any of the four subscales or on the total score (all $p > .05$). Although female students reported slightly higher mean scores across all dimensions, the differences were small and non-significant (e.g., Total: Male $M = 3.12$, Female $M = 3.25$, $U = 436.5$, $p = .297$). Kruskal-Wallis tests revealed no statistically significant differences across the four proficiency groups on any subscale or the total scale (all $p > .05$). Notably, the Comprehension Breakdown subscale showed a pattern of decreasing mean scores as proficiency increased (Elementary: $M = 3.81$; Pre-intermediate: $M = 3.45$; Intermediate: $M = 3.48$; Upper-intermediate: $M = 3.14$), though this trend did not reach significance. Pearson correlation analyses found no significant relationships between years of English study and any subscale score (all $p > .05$, $|r| < .10$), consistent with research indicating that length of study is a weaker predictor than the quality of phonological instruction (Munro & Derwing, 2015).

 DOI: *10.5281/zenodo.22078495*

Qualitative Findings (RQ2)

Analysis of the open-ended responses identified three recurring themes.

Theme 1: Cascade Breakdown.

The most pervasive pattern was the description of a cascade effect in which failure to decode a single word disrupted comprehension of subsequent utterances. One participant reported: "When I cannot decode a word, I can no longer follow the sentences after it" (P12). Another noted: "Missing one word causes a chain reaction where I miss several more" (P37). A third elaborated: "Once I miss a word, my brain keeps trying to figure out what it was, and by the time I give up, the speaker has moved on and I've lost everything in between" (P51). This pattern corroborates the quantitative dominance of the Comprehension Breakdown subscale and aligns with Field's (2008) processing bottleneck model.

Theme 2: Cognitive Overload.

Participants frequently reported that the mental effort required for phonological decoding left insufficient cognitive capacity for meaning construction. As one student articulated: "I spend so much effort decoding sounds that I lose the overall meaning" (P24). Another described the struggle by stating: "It's like my brain has to choose between hearing the sounds correctly and understanding what they mean; I can't do both at the same time" (P18). This experience of cognitive overload was particularly pronounced in connected speech contexts. These findings resonate with Cognitive Load Theory as applied to L2 listening (Vandergrift and Goh, 2012).

Theme 3: Compensatory Coping Strategies.

Despite the difficulties reported, participants drew on a range of coping strategies, including: (a) contextual guessing, which involves inferring the meaning of missed items from the surrounding context; (b) keyword monitoring, which focuses on stressed content words; (c) forward momentum, characterized by the conscious abandonment of unrecognized items; and (d) audio replay where permitted. A smaller number of students reported maladaptive responses, including disengagement and anxiety-driven paralysis. One participant acknowledged: "Sometimes when I miss too many words, I just give up

 DOI: *10.5281/zenodo.22078495*

and write any answer" (P63). These findings are consistent with Field's (2008) taxonomy of repair and expedient strategies and underscore the role of metacognitive awareness in mitigating phonological processing failure.

Table 4: Summary of Qualitative Themes

Theme	**Description**	**Illustrative response**
Cascade breakdown	Failure to decode one item triggers sequential loss of subsequent input	"Missing one word causes a chain reaction where I miss several more"
Cognitive overload	Excessive mental effort on decoding depletes resources for meaning construction	"I spend so much effort decoding sounds that I lose the overall meaning"
Compensatory coping	Strategies ranging from adaptive (contextual guessing, keyword monitoring) to maladaptive (disengagement)	Contextual guessing, keyword monitoring, forward momentum, audio replay

DISCUSSION

The present study investigated phonological decoding difficulties among Vietnamese EFL university students using a mixed methods approach. The findings offer several contributions to the understanding of bottom-up processing challenges in L2 listening comprehension.

The most prominent finding is the emergence of Comprehension Breakdown as the dimension of greatest difficulty (M = 3.45). This result suggests that for Vietnamese EFL learners, the consequences of decoding failure, rather than the initial perceptual difficulty itself, represent the most significant barrier to successful listening comprehension. This interpretation is supported by the qualitative data, in which the cascade breakdown theme dominated student accounts. The finding aligns with Field's (2008) processing bottleneck model and extends Goh's (2000) taxonomy by demonstrating that the cascade effect is

 DOI: *10.5281/zenodo.22078495*

particularly pronounced among learners from tonal, syllable-timed L1 backgrounds. The dominance of this subscale suggests that the primary locus of difficulty may not be the perceptual processing stage itself but rather the management of processing resources when perceptual demands exceed available capacity.

The strong correlation between Word Segmentation and Prosodic Processing (r = .745) suggests that these two dimensions represent overlapping facets of a shared underlying competence, namely the ability to exploit suprasegmental cues for speech parsing. This interpretation is consistent with Cutler's (2012) argument that word segmentation in stress-timed languages is fundamentally prosodic in nature. For Vietnamese learners, whose L1 employs syllable-timed rhythm, this interconnection is theoretically predictable and has practical implications for integrated curriculum design.

The relatively weaker correlation between Phoneme Perception and Comprehension Breakdown (r = .397) suggests that segmental-level confusion does not translate as directly into meaning loss as do suprasegmental processing failures. Listeners may often recover from individual phoneme misperceptions through top-down compensation, which involves using contextual knowledge to infer the intended word (Rost, 2016). In contrast, failures in word segmentation and prosodic processing may disrupt the entire parsing structure of the input, thereby making top-down compensation more difficult.

The absence of significant effects for gender, self-rated proficiency, and years of study warrants discussion. The null gender finding is consistent with the broader SLA literature (Graham, 2006). The null proficiency finding may reflect the homogeneity of the sample and limited statistical power. The finding that years of study did not predict decoding difficulty reinforces Munro and Derwing's (2015) argument that the quantity of exposure is less important than its quality, specifically the degree to which instruction incorporates explicit phonological awareness training and authentic connected speech exposure.

The qualitative findings add an important experiential dimension. The three themes, namely cascade breakdown, cognitive overload, and compensatory coping, map coherently onto the theoretical framework. Specifically, cascade breakdown corresponds to Field's (2008) processing bottleneck; cognitive overload aligns with Cognitive Load Theory (Vandergrift and Goh, 2012); and coping strategies resonate with Field's (2008) taxonomy

 DOI: *10.5281/zenodo.22078495*

of listener repair and expedient strategies. The convergence between quantitative and qualitative datasets strengthens the validity of the findings and demonstrates the value of the explanatory sequential design.

The range of coping strategies identified in the qualitative data merits particular attention. The prevalence of contextual guessing as the most commonly reported strategy suggests that Vietnamese EFL learners have developed a reliance on top-down compensation when bottom-up decoding fails. While contextual guessing can be an effective compensatory strategy in many listening situations, over-reliance on top-down processing carries risks. For instance, when the context is insufficient or misleading, learners who lack adequate bottom-up decoding skills are left without a fallback processing route (Field, 2008). Furthermore, the identification of maladaptive responses, which include complete disengagement and anxiety-driven paralysis, is particularly concerning. Such responses suggest that some learners have exhausted their compensatory repertoire and have not been equipped with metacognitive strategies necessary for managing persistent decoding failure.

An additional finding worth highlighting is the low conscious salience of connected speech features such as reduced forms and sound linking, which were rated as the least frequently experienced difficulties despite being well-documented challenges in the EFL listening literature (Cauldwell, 2013). This discrepancy between the theoretical importance of connected speech phenomena and learners' self-reported awareness of them may reflect a measurement artifact, where learners are not consciously aware of the specific phonological processes that cause their difficulties, or it may indicate that Vietnamese learners have developed partial familiarity with common reduced forms through media exposure. Future research using objective performance measures could clarify this discrepancy.

Pedagogical Implications

The findings yield several practical implications for EFL listening instruction in Vietnamese university contexts. First, given the prominence of Comprehension Breakdown, instructors should adopt a diagnostic rather than merely evaluative approach to listening assessment, investigating where and why comprehension breaks down using techniques such as pause-and-reflect protocols, partial dictation tasks, and think-aloud

 DOI: *10.5281/zenodo.22078495*

procedures that reveal decoding failures at specific points in the speech stream (Field, 2008). This diagnostic orientation can help teachers identify the specific phonological features that trigger comprehension breakdowns for their particular learner population and design remediation activities accordingly.

Second, the strong interconnection between Word Segmentation and Prosodic Processing suggests that these skills should be taught in an integrated manner rather than as isolated competencies. Activities that develop learners' awareness of English stress patterns, rhythm, and connected speech phenomena, including shadowing exercises, discourse-level dictation, and graded exposure to authentic speech at progressively increasing speeds, are likely to address both dimensions simultaneously (Cauldwell, 2013). Specifically, teachers should help learners develop the Metrical Segmentation Strategy by training them to identify stressed syllables as anchor points for word recognition, an approach that directly targets the rhythmic processing gap between Vietnamese and English.

Third, the qualitative finding that students employ both adaptive and maladaptive compensatory coping strategies points to the value of explicit metacognitive strategy instruction. Teaching learners to recognize when comprehension has broken down and to deploy appropriate repair strategies, such as contextual inference, selective attention to key words, and the deliberate decision to release unrecognized segments rather than fixating on them, can help mitigate the cascade effect that emerged as the dominant theme in this study (Vandergrift and Goh, 2012). Strategy instruction should be embedded within regular listening activities rather than taught as a separate module, so that learners develop the habit of monitoring and repairing their comprehension in real time.

Fourth, the null effect of years of study underscores the need for qualitative changes in pedagogical practice rather than simply increasing the quantity of instruction. Vietnamese EFL programs should consider incorporating systematic phonological awareness training that addresses cross-linguistic contrasts between Vietnamese and English, including minimal pair discrimination exercises, connected speech awareness activities, and stress-pattern training. The gap between classroom pronunciation models and authentic connected speech should be explicitly addressed from the earliest stages of instruction, with learners given regular and scaffolded exposure to naturalistic speech.

 DOI: *10.5281/zenodo.22078495*

Finally, the finding that even students at the intermediate and upper-intermediate levels report substantial phonological decoding difficulties suggests that these challenges are not automatically resolved through general proficiency development. Phonological processing instruction should therefore be maintained across proficiency levels, with activities increasing in complexity rather than being discontinued as learners advance.

CONCLUSION

This study has provided a multidimensional portrait of phonological decoding difficulties among Vietnamese EFL university students. Through an explanatory sequential mixed methods design combining a 28-item self-report questionnaire with open-ended qualitative responses, it has demonstrated that Comprehension Breakdown, defined as the downstream consequence of decoding failure rather than the initial perceptual difficulty itself, represents the most severely experienced dimension of phonological decoding difficulty. The finding that Word Segmentation and Prosodic Processing are tightly interconnected suggests that these dimensions share an underlying competence related to the exploitation of suprasegmental cues for speech parsing, which is a competence that is particularly challenging for learners from syllable-timed L1 backgrounds.

The absence of significant effects for gender, self-rated proficiency, and years of English study challenges the assumption that phonological decoding difficulties resolve naturally with increased exposure. Instead, these results suggest that targeted, quality-focused pedagogical intervention is required. The qualitative themes of cascade breakdown, cognitive overload, and compensatory coping provide experiential depth that enriches the quantitative findings, revealing both the psychological burden of phonological decoding failure and the range of strategies, including both adaptive and maladaptive behaviors, that learners deploy in response.

These findings contribute to the growing body of literature on bottom-up processing in L2 listening comprehension, with particular relevance to under-researched learner populations in the Asian EFL context. They offer a foundation for evidence-based pedagogical interventions that target phonological awareness, integrate prosodic and segmentation training, and develop metacognitive strategy use. It is hoped that this study will encourage further investigation into the specific phonological decoding challenges faced by

 DOI: *10.5281/zenodo.22078495*

Vietnamese and other tonal-language-background EFL learners, while informing the development of instructional approaches that build not only perceptual accuracy but also the cognitive resilience required for successful real-time listening comprehension.

ACKNOWLEDGMENTS

This study was conducted with the encouragement of the Nguyen Tat Thanh University in Ho Chi Minh City, Vietnam. My greatest appreciation goes to the Foreign Language Department employers who supported me in doing this research.

REFERENCES

1. Anderson, J. R. (1995). Cognitive psychology and its implications (4th ed.). W. H. Freeman.
2. 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: In honor of James Emil Flege (pp. 13–34). John Benjamins.
3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
4. Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), Perspectives on teaching connected speech to second language speakers (pp. 1–15). University of Hawai'i Press.
5. Cauldwell, R. (2013). Phonology for listening: Teaching the stream of speech. Speech in Action.
6. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
7. Cutler, A. (2012). Native listening: Language experience and the recognition of spoken words. MIT Press.

 DOI: *10.5281/zenodo.22078495*

8. Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. Journal of Memory and Language, 31(2), 218–236. https://doi.org/10.1016/0749-596X(92)90012-M

9. Field, J. (2008). Listening in the language classroom. Cambridge University Press.

10. Field, J. (2018). Rethinking the second language listening test: From theory to practice. Equinox Publishing.

11. 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.

12. George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference 11.0 update (4th ed.). Allyn & Bacon.

13. Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. System, 28(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

14. Goh, C. C. M. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. RELC Journal, 39(2), 188–213. https://doi.org/10.1177/0033688208092184

15. Graham, S. (2006). Listening comprehension: The learners' perspective. System, 34(2), 165–182. https://doi.org/10.1016/j.system.2005.11.001

16. Hasan, A. S. (2000). Learners' perceptions of listening comprehension problems. Language, Culture and Curriculum, 13(2), 137–153. https://doi.org/10.1080/07908310008666595

17. Le, V. C. (2019). English language teaching in Vietnam: Aspirations, realities, and challenges. In V. C. Le, T. M. H. Nguyen, T. T. M. Nguyen, & R. Barnard (Eds.), Building teacher capacity in English language teaching in Vietnam (pp. 1–20). Routledge.

 DOI: *10.5281/zenodo.22078495*

18. Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. Journal of Second Language Pronunciation, 1(1), 11–42. https://doi.org/10.1075/jslp.1.1.01mun

19. Nguyen, T. A. T. (2021). Vietnamese EFL learners' pronunciation difficulties: A cross-linguistic perspective. VNU Journal of Foreign Studies, 37(4), 142–158.

20. Nguyen, T. T. A., & Macken, M. A. (2008). Phonological transfer in Vietnamese speakers of English. In P. Bentahila, E. E. Davies, & J. Rosenhouse (Eds.), Perspectives on Arabic linguistics (pp. 207–227). John Benjamins.

21. Rost, M. (2016). Teaching and researching listening (3rd ed.). Routledge.

22. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

23. Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. Language Teaching, 40(3), 191–210. https://doi.org/10.1017/S0261444807004338

24. Vandergrift, L., & Goh, C. C. M. (2012). Teaching and learning second language listening: Metacognition in action. Routledge.

Appendix

Research Questionnaire: Phonological Decoding Difficulties in EFL Listening

Subscale 1: Phoneme Perception

1. I have difficulty distinguishing similar sounds (e.g., /l/ vs /r/, /b/ vs /p/).
2. I mishear words because I cannot identify individual sounds correctly.
3. I confuse words that sound similar (e.g., sheep/ship, live/leave).
4. Unfamiliar sound combinations make it hard for me to recognize words.
5. I miss words because their pronunciation differs from how I learned them.
6. I confuse English vowel sounds when listening (e.g., /æ/ vs /ɛ/, /ɪ/ vs /iː/).

 DOI: *10.5281/zenodo.22078495*

7. I fail to recognize words because of differences between written and spoken forms.

Subscale 2: Word Segmentation

8. I cannot tell where one word ends and the next begins in fast speech.
9. When sounds are linked (e.g., “turn off” → “tur-noff”), I fail to recognize them.
10. I miss words when sounds are dropped (e.g., “next day” → “nex day”).
11. I get confused when two words sound like one word in connected speech.
12. I fail to catch words when they are reduced (e.g., “going to” → “gonna”).
13. Fast speech makes it impossible for me to separate individual words.
14. I lose meaning when the speaker blends words together across sentence boundaries.

Subscale 3: Prosodic Processing

15. I miss unstressed words (e.g., “a, the, of, to”) because they are spoken too softly.
16. I cannot identify which words are stressed in a sentence.
17. The rhythm of English speech makes it hard for me to follow the meaning.
18. I lose track of meaning during long unstressed stretches of speech.
19. I find English stress patterns confusing because they differ from Vietnamese.
20. I miss important information because I cannot predict where stress will fall.
21. English intonation patterns confuse me when I try to understand the speaker’s meaning.

Subscale 4: Comprehension Breakdown

22. When I fail to decode a word, I can no longer follow the sentences that come after it.
23. I spend so much mental effort on decoding sounds that I lose the overall meaning.

 DOI: *10.5281/zenodo.22078495*

24. By the time I decode one word, the speaker has already moved on and I have missed what follows.
25. I cannot hold what I just heard in my head long enough to make sense of it.
26. Missing one word causes me to miss several more words in the same utterance.
27. I lose the thread of meaning when the speech is faster than I can process.
28. I understand individual words but cannot piece them together into a meaningful message.

Open-Ended Questions

Q1: When you cannot identify individual sounds or words while listening to English, what specific difficulties do you experience?

Q2: How do the rhythm and stress patterns of English affect your ability to follow what you hear?

Q3: When you fail to decode a sound or word, what happens next—and what do you do to cope?

DOI: *10.5281/zenodo.22078495*