The International Journal of Language Studies
2025, Vol. 2, No. 4, pp. 42–53
Journal page: https://ijlangstudies.org/index.php/home
DOI: https://doi.org/10.60087/ijls.v2.n4.005

ISSN 3078-2244
9 773078 224005 >

Research Article

An acoustic analysis of English-majored students' errors in pronouncing plural nouns and third-person singular simple-present verbs: A case study

Phuong Nhi Le[1]

Minh Thanh To[2]

[1] Binh Duong University, Ho Chi Minh City, Vietnam

Abstract

This quantitative study examines how 25 Vietnamese university English majors pronounce the word-final /s/ or /z/ in plurals and third-person singular simple-present verbs. Recruited via convenience and stratified samplings, the students completed three speaking tasks, producing 2,136 tokens for analysis and revealing two primary error types: dropping (74%) and mispronouncing (26%). Acoustic analysis using Praat (waveforms, spectrograms, spectral slices) interpreted that systematic mispronunciations often traceable to erroneous patterns in penultimate sounds. By providing an acoustic-phonetic diagnostic basis and concrete instructional strategies, the study empowers TESOL teachers with evidence-based practices to detect and assess word-final /s/ and /z/ errors hardly perceived by human ears, and remediate them, thereby improving accuracy and fluency in speaking English as a foreign language.

Keywords

acoustic analysis, error analysis, cognitive phonetics, classroom-applicable practices

1. Introduction

1.1 Statement of the problem

English pronunciation remains a major challenge for English as a Foreign Language (EFL) learners in Vietnam, where communicative competence is increasingly emphasized. Despite their strong grammar and vocabulary, these non-native learners of English often struggle with final consonants and consonant clusters due to limited explicit pronunciation instruction in traditional EFL teaching.

*Corresponding author: Phuong Nhi Le

Email addresses:

laclac1105@gmail.com (Phuong Nhi Le), tmthanh@bdu.edu.vn, minhthanhto@gmail.com (Minh Thanh To)

Received: 02/10/2025; **Accepted:** 05/11/2025; **Published:** 25/12/2025

Consequently, they rely on imitation without fully understanding sound production (Richard, 1976), leading to persistent errors that reduce intelligibility. That is why identifying and analysing pronunciation errors is essential for English-majored students at a university in Ho Cho Minh City—those who are expected to achieve high oral proficiency in a global context where non-native speakers outnumber native ones (Harmer, 2007).

1.2. Aim, objectives and research questions

The study reported in this paper aims to search for errors often made by second-year English-majored students at a university in Ho Cho Minh City in pronouncing the suffix -(e)s in plural nouns or third-person singular simple-present verbs either as /s/ or as /z/. Specifically, the study is supposed to:

- Identify various types of /s/ and /z/ errors made by the students in pronouncing English plural nouns and third-person singular simple-present verbs;

- Identify possible causes of these errors based on acoustic analysis;

- Propose practical suggestions to correct the errors.

To achieve its aim and objectives, the study addresses the three research questions (RQs):

RQ1. What are main types of errors that the students make in pronouncing the suffix -(e)s as /s/ or /z/ in plural nouns or third-person singular simple-present verbs?

RQ2. What are possible acoustic causes of these errors?

RQ3. What are optimal suggestions to correct the errors?

2. Literature Review

2.1. Theoretical framework

2.1.1. Definitions of some key terms (To, 2025a; To & Phan, 2025b)

- *The English suffix -(e)s* is attached either to a noun to indicate its plurality (*cat* → *cats*, *bus* → *buses*) or to a verb to indicate its third-person singular simple present (*walk* → *walks*, *apply* → *applies*).

- *English oral stops*: “The airflow is completely blocked in the oral cavity for a brief period because the velum is raised to shut off the nasal cavity and, at the same time, either the lips are pressed together, or the tongue touches some part of the roof of the mouth to shut off the oral cavity. the complete closure is then suddenly released, and the airflow escapes with an explosive sound.” (To & Phan, 2025b, p.16) Oral stops, such as /p/, /t/, and /k/, are characterized by a closure phase, a hold (or occlusion) phase, and a release phase, and they do not allow nasal airflow because the velum is raised during the closure.

- *Voiced* vs. *voiceless*: “The airflow from the lungs moves up through the trachea (also called windpipe) and through the opening between the vocal cords, which is called the glottis. If the vocal cords are apart so that the airflow goes freely through the glottis into the oral cavity, then sounds are voiceless: /p/ and /s/ in *super* /ˈsjuːpə/ or /ˈsuːpər/. If the vocal cords are close to each other, the airflow forces its way through and causes them to vibrate, resulting in voiced sounds, like /b/ and /z/ in *buzz* /bʌz/.” (To & Phan, 2025b, pp.10–11)

2.1.2. Acoustic phonetics

Acoustic phonetics focuses on the physical characteristics of speech as it is transmitted through sound waves in the atmosphere (Yule, 2020). Acoustic analysis is particularly valuable for the study of speech for various reasons, one of the most significant being its ability to connect the processes of speech production and speech perception through the acoustic signal.

Therefore, acoustic analysis provides insights into both the behaviours of the speaker and the perception of the signal by the listener. Acoustic analysis complements studies of speech physiology and speech perception, as it captures the perturbations in the air caused by speech, which are then detected by the ear and converted into electrical signals for processing in the brain, including the identification of meaningful linguistic units such as sounds, words, and sentences (Kent & Kim, 2008).

2.1.3. Acoustic measurements and analyses of /s/ and /z/

Articulatorily, though both are alveolar fricatives, /s/ is voiceless while /z/ is voiced (Yule, 2020). Acoustically, this contrast is represented by a low-frequency voicing bar and periodic energy in /z/ versus the aperiodic, high-frequency noise of /s/.

Acoustic analysis was conducted using waveforms and spectrograms generated in Praat. Waveforms plot amplitude (loudness) against time, allowing for the visual identification of periodicity. Regular vibration patterns indicate voiced sounds, while irregular patterns indicate voiceless sounds (Davenport & Hannahs, 2020).

Spectrograms provide a more detailed visualization by plotting frequency against time, with intensity represented by the darkness of the display. Wide-band spectrograms, which offer high temporal resolution, were primarily used in this study to analyse consonantal events.

According to Ladefoged & Johnson (2015), the features examined in the spectrograms are:

①Formants (F): These appear as dark bands (F1, F2, F3) and indicate the resonant frequencies of the vocal tract. While most salient for vowels, their transitions provide crucial cues for a consonant's manner (related to F1) and place of articulation (related to F2 and F3).

②Fundamental frequency (F0): This is the lowest frequency in a voiced sound, corresponding to the rate of vocal fold vibration. The presence of a low-frequency F0 band (around 200-300 Hz) and its harmonics beneath higher-frequency noise is a primary acoustic indicator of voicing bar, essential for distinguishing between phonemes like /s/ and /z/.

③Harmonics: When the vocal folds vibrate, they produce what are called harmonics of their fundamental frequency of vibration. Harmonics are vibrations at whole-number multiples of the fundamental frequency. Thus, when the vocal folds are vibrating at 100 Hz, they produce the first harmonics at 100 Hz, the second at 200 Hz, the third at 300 Hz, and so on.

④Aperiodic noise: The wave pattern is irregular and non-repeating over time, simply understood as random noise, in contrast to the regular, repeating pattern of vibration, known as periodic sounds (like vowels or voiced consonants).

2.1.4. Error analysis

Error analysis (EA), first introduced by Corder's influential paper on "the meaning of learner errors" (1967), is a pioneering method in *second language learning* (SLA) research whereby actual learner errors in a second language should not be seen as "bad habits" to be eliminated, but as valuable insights into the learning process.

Brown (1980) defines the concept of EA as the process of observing, analyzing, and categorizing the systematic deviations from the rules in the second language, thereby revealing the systems operated by the learner. This definition further suggests that EA is a strategy for identifying, classifying, and systematically interpreting the erroneous forms produced by foreign language learners, theorized in conjunction with linguistic principles and procedures.

2.2. Previous studies on problems facing EFL learners in pronouncing English word-final consonants and consonant clusters

Internationally, Abker (2020), in a study of Saudi EFL students, found significant difficulty in correctly pronouncing English morphemes for plural nouns and verb inflections, attributing errors to a lack of practice and inattention to morphological rules.

Back in Vietnam, Nguyen (2008) focused on the interlanguage phonology of advanced Vietnamese learners of English, specifically their production of two-member final consonant clusters (2MFCs). This research identified the most difficult clusters, catalogued common modification tactics used by learners, and examined how task type influences pronunciation output. Nguyen (2019) investigated the intelligibility of word-final consonants produced by Vietnamese learners of English. Using production tasks (word lists, text reading) and perception tests evaluated by native and non-native judges, the study adapted frameworks from Nguyen & Brouha (1998) and Sato (1984) to analyze errors and propose pedagogical solutions.

Riaño (2021) examined the transfer of Northern Vietnamese phonetic features to English, focusing on consonant clusters and voiceless final obstruents using auditory and acoustic methods. Similarly, Tran & Nguyen (2022) detailed EFL Vietnamese learners' errors in producing English consonant clusters, especially sound substitutions and omissions, advocating for explicit instruction. In the same vein, Slowik & Dung (2022) explored the pronunciation patterns of South Vietnam non-native speakers of English, analyzing challenges with vowels, consonant clusters and stress patterns to inform targeted teaching strategies. At the same period, Lam and Thi (2022) conveyed a quantitative study, investigated pronunciation errors in English consonant clusters among 39 Vietnamese EFL university learners. Using a pronunciation test, it found that mistakes varied by cluster type, with the highest mispronunciation occurring in clusters containing voiceless plosives. Learners also frequently simplified three-consonant clusters by deleting one or more consonants. The study concludes with pedagogical implications for teaching and learning English pronunciation in the Vietnamese context.

2.3. Research gaps

Firstly, prior studies have primarily relied on manual detection and evaluation of pronunciation errors using human auditory judgment. Even when speech was slowed down with technological assistance, researchers were unable to pinpoint the exact phonetic phase responsible for the error. This lack of precision means that the root cause of mispronunciation often remains unidentified. By contrast, employing acoustic analysis as a tool for examining sound production results in precisely isolating **the** phase within a sound—at the millisecond level—that triggers the problem. This enables the classification of recurring acoustic error patterns, thereby offering a more systematic understanding of pronunciation difficulties.

In addition, prior research has often drawn conclusions based primarily on such measures as pre-tests, post-tests, and interviews dataset. While useful, these measures provided limited insight into learners' actual training processes, study habits, and attitudes toward pronunciation practice. Recognizing this limitation, the current study aims to deliver a more concrete and detailed body of evidence.

By analyzing the mechanisms of frequency, formants, and the presence or absence of voicing, this study revealed specific mispronunciation habits at a fine-grained level, offering a richer understanding of how the students approached and internalized pronunciation.

As for pedagogical suggestions, prior studies have tended to propose exercises, e.g., focusing on word-final consonant clusters, without identifying the precise segment where learners' errors occur. Yet even a very brief phase within a sound—the release burst of the voiceless alveolar oral stop /t/—can significantly affect the voicing of subsequent sounds. To address this gap, the current study suggested solutions enhanced by AI applications that helped learners DETECT mispronunciations at an extremely detailed level. Also, it proposed targeted exercises for specific errors, such as the omission of the word-final /s/ or /z/. This type of error reflected learners' perception of the suffix *-(e)s*. By designing exercises that activate learners' cognitive awareness while preserving their fluency, the study raised effective pedagogical interventions.

Collectively, prior studies established a research gap: while the difficulty of word-final consonants for Vietnamese learners is well-documented, there is a need for a focused acoustic analysis that directly links specific phases of final consonants that are considered problematic, leading to the mispronunciation of /s/ and /z/. The detailed analysis reported in this paper provides deeper insight into the root causes of the very mispronunciation, suggesting targeted practices that address the issue at the level of learner awareness and conscious control.

3. Methodology

3.1. Research design

This quantitative study relied totally on statistical data set coded from the results from a three-task speaking experiment conducted by 25 English majors in their fourth semester at a university in Vietnam. Employed first was convenient sampling and then stratified sampling. This was regarded as suitable and sufficiently reliable because it minimized random personal variations and increased the consistency of the errors collected, thus creating a more comparable dataset for error analysis. Errors were identified based on the adapted model (see *Figure 3.2*) and then categorized into two main error types: 1. Dropping the word-final /s/ or /z/ and 2. Mispronouncing the word-final /s/ or /z/. Acoustic analysis, using such tools as waveforms, spectrograms and spectral slices, provided an objective method for examining the subtle pronunciation word-final contrasts of /s/ and /z/. Such an approach is essential for reliably identifying phonetic deviations in pronunciation research (Kent & Kim, 2008).

3.2. Sample and sampling

Convenient samples were 69 English sophomores who took English phonology and morphology courses already, but instructor observations indicated they persistently made basic phonological errors, particularly in producing the word-final /s/ or /z/. They all did the reading task and joined individual interviews, but only 50 students remained after audio noise filtering. Background noise was removed without editing speech.

Stratified sampling then selected odd-numbered cases from the 50, yielding 3,550 reading tokens (71×50) and 722 interview tokens (~10×50); halved for the article to **2,136 tokens**, which was quite acceptable.

3.3. Research instruments

3.3.1. Technical tools

Applied as the study's technical tools are ①*An external microphone* used to capture high-quality audio in order to ensure clarity for acoustic analysis while minimizing participant awareness of the recording process (Martimo, 2015); ②*Acoustic Analysis of Consonants* (AAC) (Martimo, 2015) used to conduct quantitative analysis; and ③*Praat software* (Styler, 2013) used to process recordings for detailed phonetic examination.

3.3.2. Error identification model

Adapted from Corder's (1971) framework for second-language utterances to analyze the two word-final sounds /s/ and /z/ (see *Figure 3.1* and *Figure 3.2*), *Error Identification Model* categorizes two general pronunciation error types: (i) *dropping* (complete omission of the target sound) and (ii) *mispronouncing* (production of an incorrect sound).

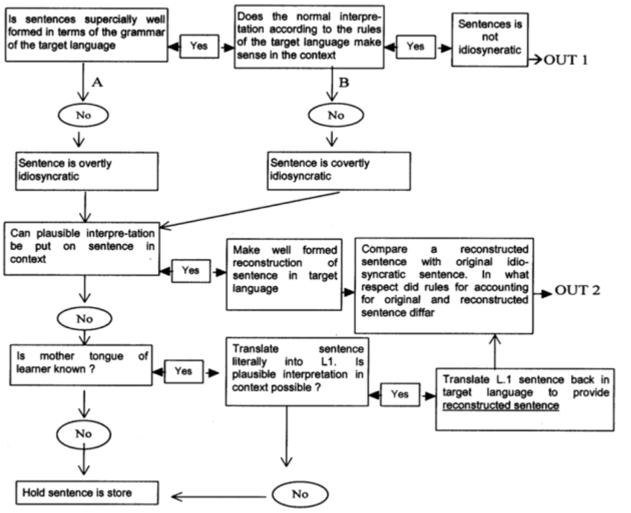

Is sentences supercially well formed in terms of the grammar of the target language
Yes
Does the normal interpre-tation according to the rules of the target language make sense in the context
Yes
Sentences is not idiosyneratic
OUT 1
A
B
No
No
Sentence is overtly idiosyncratic
Sentence is covertly idiosyncratic
Can plausible interpre-tation be put on sentence in context
Yes
Make well formed reconstruction of sentence in target language
Compare a reconstructed sentence with original idio-syncratic sentence. In what respect did rules for accounting for original and reconstructed sentence diffar
OUT 2
No
Is mother tongue of learner known ?
Yes
Translate sentence literally into L1. Is plausible interpretation in context possible ?
Yes
Translate L.1 sentence back in target language to provide reconstructed sentence
No
Hold sentence is store
No

Figure 3.1. Model for identifying erroneous or idiosyncratic utterances in a second language, (Corder, 1971)

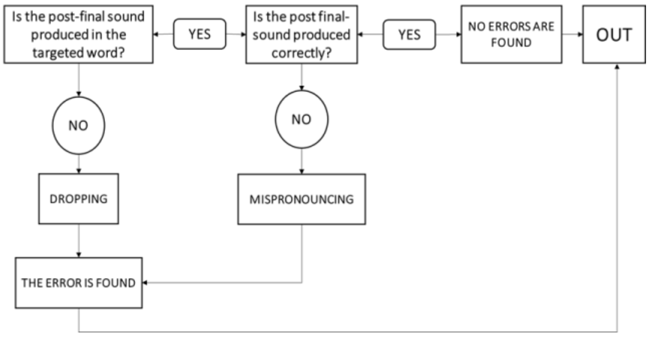
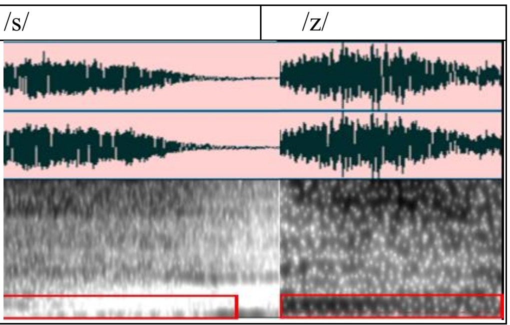
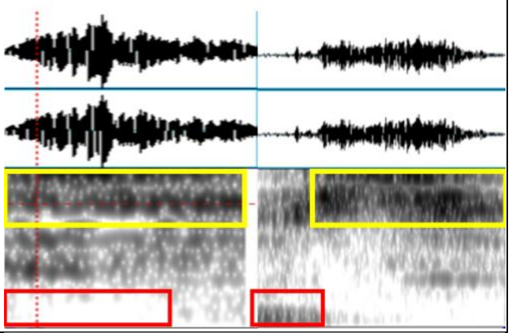
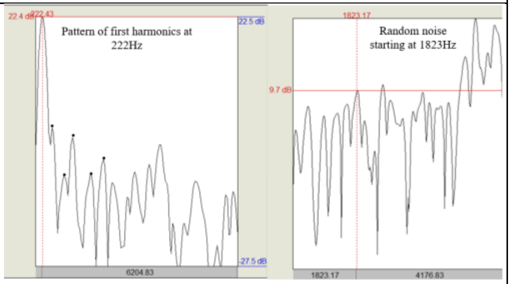

Is the post-final sound produced in the targeted word?
YES
Is the post final-sound produced correctly?
YES
NO ERRORS ARE FOUND
OUT
NO
NO
DROPPING
MISPRONOUNCING
THE ERROR IS FOUND

Figure 3.2. Model for identifying erroneous pronunciation in the word-final /s/ or /z/

3.3.3. Acoustic criteria for error identification

Spectrographic analysis was used to identify errors objectively. The relative amplitude of harmonics (indicated by the darkness of bands on a spectrogram) was the primary diagnostic (Lieberman & Blumstein, 1988). Below are the key indicators for the English fricatives /s/ and /z/:

	/s/	**/z/**
Voicing bar (spectrogram)	No periodic vocal fold vibration (voicing bar)	Periodic vocal fold vibration (voicing bar) in low frequency (about 200~350Hz)
Aperiodic noise (spectrogram)	Blocks of highest-frequency random noise (about 3500~5500Hz) **beginning** <u>right at the onset</u> of the /s/ sound.	Blocks of highest-frequency noise (random noise) (about 3500~5500Hz) **beginning** immediately <u>after the voicing bar ends</u> (because vocal-fold vibration and random noise cannot occur at the same time)
First harmonics (spectral slice)	Absence of first harmonics + big blocks of energy at the highest frequency 3500~5000 Hz)	Pattern of first harmonics (200~350Hz) + big blocks of energy at the highest frequency 3500~5000 Hz)

	/s/	/z/
Voicing bar (spectrogram)		
Aperiodic noise (spectrogram)		
First harmonics (spectral slice)		

3.4. Data collection procedures

Designed with the support of electret condenser microphones to collect 2,136 recorded audio files from 25 students' pronouncing /s/ and /z/ was a <u>*three-task*</u> speaking experiment:

Task 1. <u>*Paragraph reading*</u> (controlled): Participants read two paragraphs from an elementary-level textbook;

Task 2. <u>*Word list reading*</u> (less controlled): Participants read a two-column word, all of which appeared in the preceding paragraphs;

Task 3. <u>*Semi-structured interview*</u> (spontaneous): Participants joined an open-ended conversation that elicited unscripted speech.

To minimize the students' metalinguistic awareness and capture their pronunciation habits, Task 2 was deliberately placed after Task 1 (Tim & Paul, 2008; Labov, 1972; Trudgill, 1974; Martimo, 2015), though both ensured consistent token production and style shifting observation.

3.5. Data analysis procedures

3.5.1. Corpus processing

First, the researcher and the volunteer native speaker (called "the native" from now on for short) manually verified **text** and **phonemic transcriptions** transcribed from interview audios respectively by Otter.ai speech-to-text (Liang & Fu, 2020) and by tophonics.com (Hirch, 2016).

Next, the researcher prepared **the textual corpus** for alignment with **the acoustic data** by:

- Filtering out the background noise by Praat;

- Extracting each individual target word, either a plural noun or a third-person singular simple-present verb, from the continuous speech;

- Segmenting utterances, from the unscripted interviews (Task 3), based on perceptible pauses (Roach, 2009);

- Noting phonetic phenomena like connected speech at word boundaries.

Finally, the researcher categorized **the extracted tokens** by speaking style (scripted reading vs. unscripted interview). All 2,136 tokens were exported as **individual .wav files**, together with **corresponding Praat-collection files** containing phonetic markers and spectrographic data.

3.5.2. Error identification and acoustic analysis in Praat

Initially, the number tokens collected through the experiment were identified and categorized basing on the *Model of identifying erroneous pronunciation in the word-final* /s/ or /z/ then coded into Excel sheet (see *3.5.3*).

All 2,136 tokenized .wav files were analyzed in Praat. Each participant's production was contrasted against native-speaker models. The analysis focused on the word-final consonants, examining spectrograms for the presence/absence and voicing quality of /s/ or /z/.

This multi-faceted acoustic inspection formed the basis for objective generalizing error patterns.

3.5.3. Coding scheme

Classifying each token was a *two-tier* numerical coding system:

● *Word-final error*: 0 = Correct production; 1 = Dropping (omission). 2 = Mispronouncing;

● *Final segment influence*: For tokens coded as 2, the effect on the preceding final consonant was noted as 0 = Deletion of the final; 1 = Substitution of the final. 2 = Mispronunciation of the final.

Each token thus received a composite code (e.g., 2-1 for a mispronounced word-final sound with substitution of the preceding consonant). All codes were logged in a master Excel spreadsheet for quantitative analysis.

4. Results and Discussion

4.1. Analysis of data

4.1.1. Overall error occurrence

Table 4.1. Occurrences and percentages of non-errors and errors of the three tasks

Categories	**Occurrences**	**Percentages**
Non-errors	1208	57%
Errors	928	43%
Total	2136	100%

Table 4.1 shows that the analysis of 2,136 tokens confirmed **a substantial overall error rate** of 43%, with 57% of correct products. This is considered statistically significant for it highlights the prevalence of errors in the students' spoken output.

Table 4.2. Occurrences and percentages of non-errors and errors of Task 1, Task 2, and Task 3

Categories	**Task 1**		**Task 2**		**Task 3**	
	Occurrences	**Percentages**	**Occurrences**		**Occurrences**	**Percentages**
Non-errors	560	57%	574	72%	74	20%
Errors	415	43%	226	28%	287	80%
Total	975	100%	800	100%	361	100%

As shown in *Table 4.2*, a breakdown by task reveals a definitive pattern linked to 3 different speaking tasks: A moderate error rate (43%) for Task 1. Paragraph reading; the lowest error rate (28%) for Task 2. Word list reading; the highest error rate (80%) for Task 3. Unscripted interviews. The increase, where errors nearly triple from the most controlled to the most spontaneous task, demonstrates that controlment and spontaneity in the tasks directly influences phonological accuracy. Crucially, the considerable error rates in the controlled reading tasks (28% and 43%) cannot be attributed solely to the pressure of spontaneity. This suggests that errors are also rooted in language habits, which surface even when participants have full visual support and no time constraints.

4.1.2. Dropping errors and Mispronouncing errors

Table 4.3. Occurrences and percentages of dropping and mispronouncing errors in Task 1, Task 2 and Task 3

CATEGORIES	**TASK 1**		**TASK 2**		**TASK 3**	
	Occurrence N	Percent %	Occurrence N	Percent %	Occurrence N	Percent %
Dropping	264	63,6%	144	63,7%	277	96,5%
Mispronouncing	151	36,4%	82	36,3%	10	3,5%
Total	415	100,0 %	226	100,0 %	287	100,0 %

Table 4.3. shows that the percentages of the error type in the two tasks stay equally similar: 63,6% in Task 1 and 63,7% in Task 2, both being done without any speech rate control or time constraint. In other words, even if the students were

exposed to **the same target words** twice, enough to trigger their awareness, they still made the same number of mistakes in the two tasks. This indicates that (i) **text types** and **forms of tasks** did not cause any changes in the proportions of dropping errors, and that (ii) force of language habit formation of sound clusters results in omitting the suffix *-(e)s* in the two reading tasks.

In Task 3, there is a significant change of the proportion of this error type. It dramatically surges 96.5%. This suggests distinguishable speech types (reading and unscripted speaking) provoke dissimilar levels of dropping errors. Also, **dropping** is almost as 27.7 times as **mispronouncing** (277 dropping errors: 10 mispronouncing errors), which means **almost all the errors belong to the dropping type**. As described, Task 3 is unscripted and totally free speaking, the test-takers would answer questions in the interview, the answers were unprepared, this is most like English communicative conversations. This demonstrates that **the level of discursive freedom** directly influences **phonological accuracy**.

As a result, the feasible reasons regarding dropping errors are concluded as follows:

- Different speech types (reading and unscripted speaking) results in dissimilar levels of dropping errors in Task 1, Task 2 and Task 3;
- Habit of producing sound cluster results in omitting the suffix *-(e)s* in reading Task 1 and Task 2 even with the same targeted words;
- The lowest level of discursive freedom in speech intentionally causes the most dropping word-final sounds in plural nouns or third-person singular simple-present verbs of the tasks.

Table 4.3. also shows that Mispronunciation (incorrect production of the target sound) accounted for 36,4% and 36,3% respectively for Task 1 and Task 2. Unlike dropping, these numbers suggest the students' some awareness of the required suffix, albeit with faulty execution.

Notably, mispronunciation rates were also similar in Task 1 and Task 2 but plummeted to just 3,5% in Task 3. This number does not imply the participants are fully aware of pronouncing the /s/ and /z/ precisely because the other 96.5% goes to dropping errors. The significant low percentage during the interview underscores the particular difficulty of maintaining the suffix /s/ or /z/ in real-time communication.

4.1.3. Frequency distribution description

A frequency distribution analysis identified the specific words most prone to mispronunciation. For example, in Task 1, the word *gets* was the most problematic, accounting for 14.4% of mispronounced /s/ tokens (*Figure 4.1*). Similarly, descriptive statistics detailing the distribution for /s/ and /z/ in all tasks highlight the students' consistent trouble spots.

Figure 4.1. Mispronouncing /s/ in the targeted words in Task 1

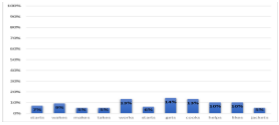


Figure 4.2. Mispronouncing /z/ in the targeted words in Task 1

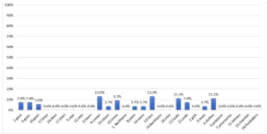


Figure 4.3. Mispronouncing /s/ in the targeted words in Task 2

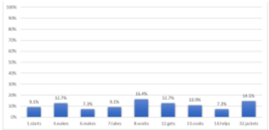


Figure 4.4. Mispronouncing /z/ in the targeted words in Task 2

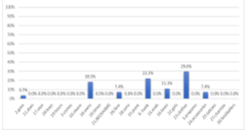
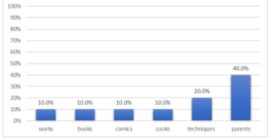


Figure 4.5. Mispronouncing /s/ in the targeted words in Task 3



4.1.4. Key results

Below are some key results from the above data analysis:

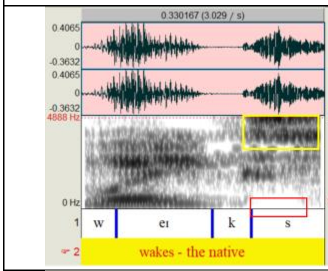
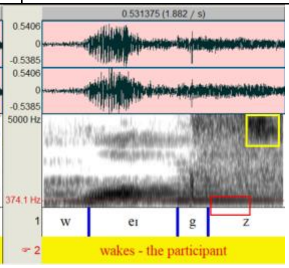
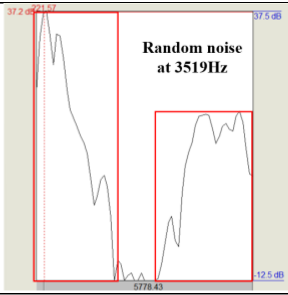
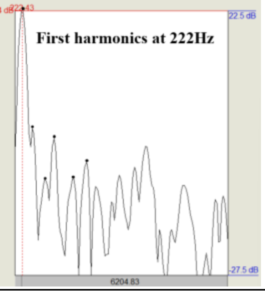
1. Determinative speech style: Spontaneous speech (Task 3) leads to a significantly higher rate of omission (96.5% dropping) compared to scripted reading (~64%).
2. Habitual omission: The consistent error rate across two different scripted tasks suggests omission is an entrenched production habit, not a task-specific effect.
3. Targeted difficulties: Frequency analysis reveals a number of words (e.g., ***gets***) were systematically more challenging, providing concrete targets for pedagogical intervention.

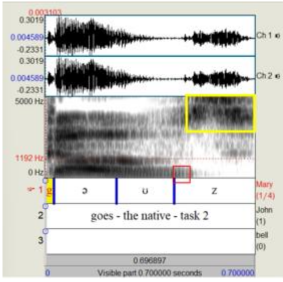
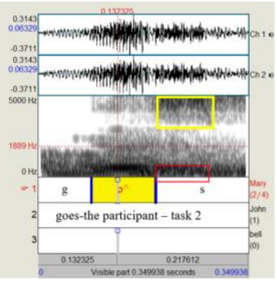
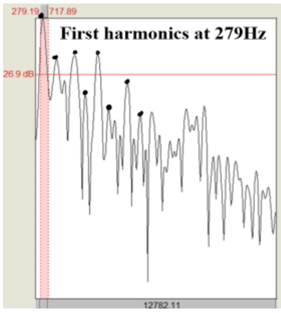
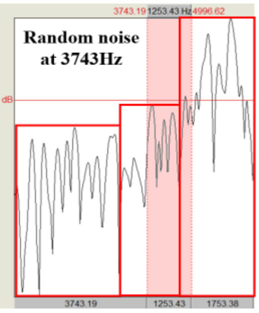
4.2. Discussion of Results

4.2.1. Acoustic analysis

Intentionally used in this part is the AAC in Praat software because such features as formants, spectral slice, and sound waves are crystal clear to analyze the distinct features of /s/ and /z/ (see *3.3.3*).

Acoustic cues of the correct /s/	*Acoustic cues of the mispronounced /s/*

	detected as /z/
	
	
• There is no vertical striation (as squared red in the spectrogram), indicating no vibration of the vocal folds which means no voicing. • There are big blocks of energy concentrated at the highest frequency (from 3519Hz to 5000Hz) (as squared yellow in the spectrogram), starting right at the onset (beginning) of the sound, corresponding the /s/ sound. • It does not contain a systemic pattern of first harmonics (as squared red in spectral slices) but two blocks of random noise at 3519Hz.	• Voicing bar formed by series of vertical striations appears obviously at onset (beginning) of the sound (as squared red in the spectrogram), signifying vibration of the vocal folds, which happens a voicing bar. • The big blocks of energy at the highest-frequency (from 3500Hz to 5000Hz) start right after the voicing bar ends (as squared yellow in the spectrogram), corresponding the /z/ sound. • There happens a systemic pattern of first harmonics (at 222Hz) (the black dots) at the onset of the sound in spectral slices.

Acoustic cues of the correct /z/	*Acoustic cues of the mispronounced /z/ detected as /s/*
	
	
• There are vertical striations (as squared red in the spectrogram), indicating vibration of the vocal folds which happens a voicing bar • The big blocks of energy at the highest-frequency (from 3500Hz to 5000Hz) (as squared yellow in the spectrogram) start right after the voicing bar ends, corresponding the /z/ sound. • There happens a systemic pattern of first harmonics (at 279Hz) (the black dots) at the onset of the sound in spectral slices.	• Voicing bar formed by series of vertical striations (as squared red in spectrogram) appears obviously at onset (beginning) of the sound, signifying vibration of the vocal folds, which happens a voicing bar. • There are big blocks of energy concentrated at the highest frequency (from 3743Hz to 5000Hz), starting right at the onset (beginning) of the sound, corresponding the /s/ sound. • It does not contain a systemic pattern of first harmonics in spectral slices but 2 blocks of random noise at 3743Hz.

However, when the voiceless word-final /s/ is wrongly pronounced to the voiced /z/, it is detected as /z/. Therefore, there is not much in investigating the errors themselves. In fact, actual implied reasons hide in erroneous patterns of word-final consonants or its penultimate consonants and vowels.

4.2.2. Erroneous patterns in the penultimate consonants and vowels of /s/ or /z

Below are patterns of the penultimate sounds of /s/ and /z/ that contribute to their mispronunciation.

Pattern 1. Absence of the release burst[1]

Take, for example, **in the voiceless** /k/**: [–release burst], [+voicing]**

In contrast to the native's clear and high-frequency release burst (as yellow-squared in *Image 4.1a*) (1-5kHz) and absence of voicing in both /k/ and /s/ (as red-squared in *Image 4.1a*), the student's producing the voiceless /k/ (e.g., in ***works***) (see *Image 4.1b*) showed (i) no voiceless release burst of /k/, and (ii) a continuous voicing bar at ~250-270Hz (as red-squared) extending from the preceding vowel /ɜː/. This on-going voicing resonated through the voiced /r/ without the blockage of the release burst, causing the intendedly voiceless /s/ to sound just like its voiced counterpart /z/.

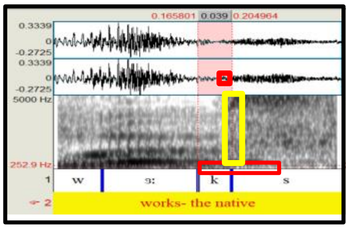
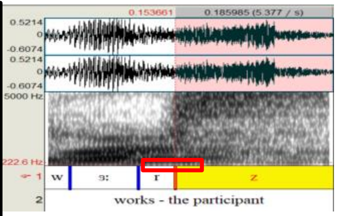
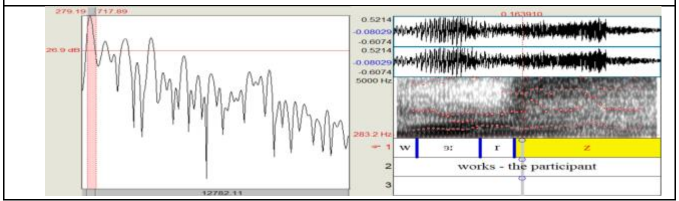
Image 4.1a. The native's clear and measurable release burst of /k/	*Image 4.1b. The student's voicing bar in the intendedly voiceless* /s/
	

Image 4.2. The student's first harmonics in the intendedly voiceless /s/

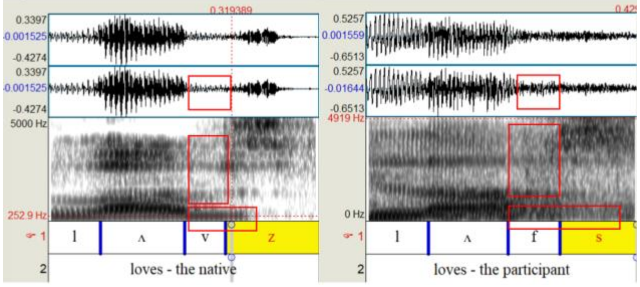


Pattern 2. Absence of the voicing

Take, for example, **in the voiced** /v/**: [–voicing], [+high-frequency energy block]**

In *Image 4.3*, compared to the native's clear and measurable blocks of high-frequency energy from 2700Hz to 4000Hz and a voicing bar at 253Hz, the student's producing the voiced /v/ (e.g., in ***loves***) showed (i) no voicing bar at low-frequency (~250 Hz), and (ii) high-frequency blocks of energy around 3,000 to 4,000 Hz, causing the intendedly the voiced /v/ to sound just like its voiceless counterpart /f/.

Image 4.3. The native's clear voicing bar and measurable blocks of high-frequency energy of /v/
vs. *the student's lack of voicing bar and energy blocks of* /v/

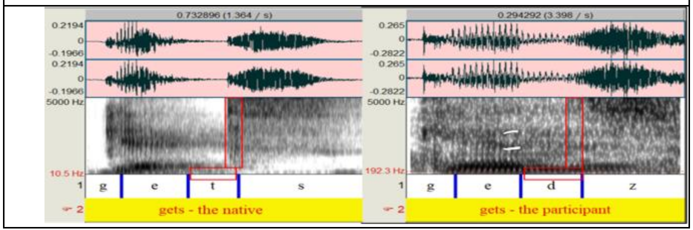


Pattern 3. Presence of the voicing

Take, for example, **to the voiceless** /t/**: [+formant transition**[2]**], [+maintaining F2 and raising F3]**

In *Image 4.4*, compared to the native's clear burst but no voicing bar, the student's producing the voiceless /t/ (e.g., in ***gets***) showed (i) a distinct voicing bar at ~192.3Hz throughout the duration of the intendedly voiceless /t/; and (ii) the visible formant transitions from the preceding vowel /e/, with F2 and F3 maintained in parallel, causing the intendedly voiceless /t/ to sound just like its voiced counterpart /d/.

Image 4.4. The native's producing /g/, /e/, /t/, *and* /s/, *with clear release burst in* /t/
vs. *the student's producing* /g/, /e/, /d/, *and* /z/ without release burst in /t/



Pattern 4. Absence of the glide

Take, for example, **in the diphthong /əʊ/: [+vowel], [–gliding]**

In *Image 4.5*, compared to the native's unchanging formants (F1~587Hz, F2~1900Hz, F3~2400Hz) with total

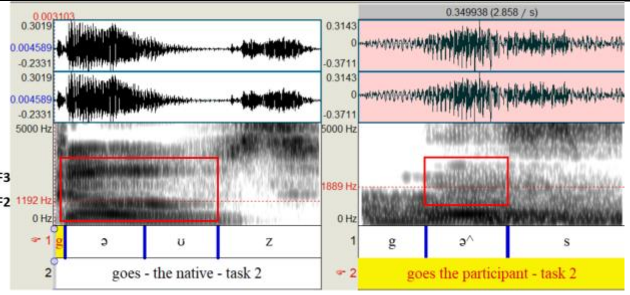
[1] The **release burst** a **very brief (10-30 ms)**, **aperiodic (noisy) spike of energy** across a wide range of frequencies, resulting from the sudden release of the built-up pressure behind a complete closure in the vocal tract (Ladefoged & Johnson, 2015).

[2] **Formant transitions** are the rapid changes in vowel resonance frequencies (formants) that occur when the vocal tract moves between consonants and vowels; acoustically they appear as short sweeps in the first two or three formant tracks on a spectrogram and signal articulatory movement and place of articulation (Ladefoged & Johnson, 2015).

0.71s in duration, indicating the on-going glide from /ə/ and /ʊ/, the student's producing the diphthong /əʊ/ (e.g., in ***goes***) showed no glide of F2 and F3 from /ə/ to /ʊ/. This created a weak vowel that was unable to keep vibration from the diphthong /əʊ/ to the consonant /z/, causing the intendedly voiced /z/ to sound just like its voiceless counterpart /s/.

Image 4.5. The native's clear glide transition of F2 and F3 in /əʊ/

vs. *the student's lack of glide transition of F2 and F3 in* /əʊ/



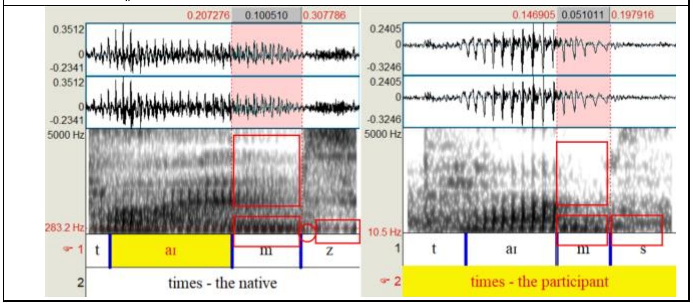
Pattern 5. Failure in maintaining the voicing

Take, for example, **in the voiced /m/: [–long duration]**, **[–weak formant range]**

In *Image 4.6*, compared to the native's measurable weak formant range in the nose area and a strong and long voicing in /m/, the student's producing a qualified voiced /m/ (e.g., in ***times***) showed (i) a rather short duration of the voiced /m/ (0.051s by the student vs. 0.1s by the native); (ii) and no typical pattern for the voiced /m/ (e.g., the weak frequency ranges around 1000Hz and 3000Hz where the sound's energy drops). The unmaintained formant range fails to act as a voice source, causing the intendedly voiced /z/ to sound just like its voiceless counterpart /s/.

Image 4.6. The native's measurable weak formant range and long duration of /m/

vs. *the student's lack of weak formant range and long duration of* /m/



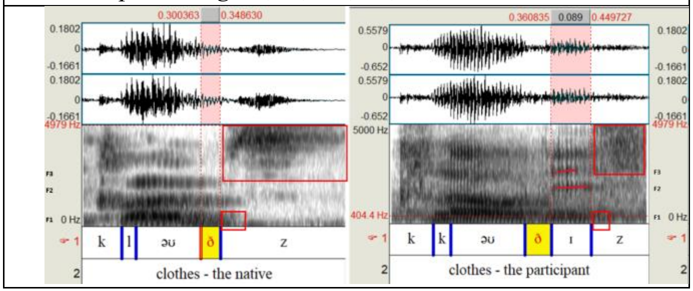
Pattern 6. Insertion of vowel formants

Take, for example, **to the vowel /ɪ/ before the voiced /z/ to rhyme a new syllable: [+vowel formants]**

The student inserted a short, high-front vowel /ɪ/ between the final consonant and the voiced /z/, creating an additional syllable (e.g., /-dz/ → /-dɪz/).

In *Image 4.7*, compared to the native's clear voicing bar and a block of the highest-frequency energy, indicating the voiced /z/, right after /ð/, the student's production showed a clear formant structure identifying the vowel: F1 at ~400Hz (high tongue body), F2 at ~1920Hz (front vowel), and F3 at ~2560Hz (slight lip spreading). This inserted vowel rhymed a new syllable.

Image 4.7.The native's producing /k/, /l/, /əʊ, /ð/, /z/vs.*the student's producing* /k/, /l/, /əʊ, /ð/, /ɪ/, /z/



5. Conclusion

5.1. Answers to the three research questions

The study's findings satisfactorily answered its three research questions.

Answer to RQ1: The two main types of errors that the students made in pronouncing the suffix *-(e)s* in plural nouns or third-person singular simple-present verbs are either **dropping** or **mispronouncing** /s/ or /z/.

Answer to RQ2:

● As for **dropping /s/ or /z/**, there are three possible causes: (1) language habit of producing sound clusters with the word-final /s/ and /z/, (2) different speech styles, and (3) levels of discursive freedom.

● As for **mispronouncing /s/**, there are five detailed acoustic patterns: (4) the first three concerning the **absent** release burst in the penultimate /p/, /t/, or /k/; (5) the last two concerning the **added** voicing to the penultimate /t/ or /k/;

● As for **mispronouncing /z/**, there are five detailed acoustic patterns: (6) the **absent** voicing in the penultimate /d/ and /m/; (7) the lack of gliding in the penultimate /əʊ/; (9) the absent voicing bar in the penultimate /v/; (10) the inserted vowel formants before /z/ to rhyme a new syllable.

Answer to RQ3:

To improve dropping /s/ or /z/:

The habitual omission of word-final sounds requires activities that force learners' awareness and production of the suffix *-(e)s*. A structured text-based exercise with steps is recommended: ①selecting a text rich in plural nouns or third-person singular simple-present verbs; ②having learners read these texts aloud and re-read it to activate awareness of the suffix *-(e)s*; ③ designing comprehension questions that specifically target these forms (e.g., after the sentence "Noel wakes up at 4 am," the question "What does Noel do at 4 am?" requires the response "wakes up," ensuring production of the target /s/).

To improve mispronouncing /s/ or /z/:

- Practice individually: Drill troublesome final clusters (e.g., /ts/, /lps/, /ndz/, /vz/) individually, focusing on clarity before increasing speed for fluency.
- Use targeted tongue twisters: Employ tongue twisters saturated with challenging final consonants (/p, t, k, m, n, d, v/) to build muscular habit and automaticity.
- Contrast with minimal pairs: Practice minimal pairs (e.g., ***starts*** vs. ***stars***, ***loves*** vs. ***laughs***) at word, sentence, and paragraph levels. Use voice recognition tools (e.g., Google Speech-to-Text) to verify accurate production and perception.
- Compare diphthongs acoustically: Use Praat to visually compare the formant transitions of the English diphthong /əʊ/ with the nearest Vietnamese counterpart (e.g., /âu/), highlighting critical differences in the glide.
- Use AI platforms: For more practices, learners can use AI platforms to generate conversations focused on plural nouns or third-person singular simple-present verbs. The AI can then provide constructive feedback at the sound-segment level and on overall fluency.
- Practice in pairs or groups: During pair or group practice, participants should aim to reduce errors in casual speech while maintaining a balance between fluency and accuracy. As English majors, developing a high degree of conscious awareness in speaking is essential to meet the expected language competency.

5.2. Significance

The study's findings are significant because mispronouncing the analysed word-final sounds hinders high-level spoken English competency. This is particularly crucial because these sounds are among English's most frequent consonants. Sounds such as /m/, /n/, /ŋ/ constitute 18.45% of all consonants, with /n/ being the most frequent overall (Mines, Hanson, & Shoup, 1978). Hence, sounds like /p/, /t/, and /k/ account for 29.21% of consonants. Collectively, aforementioned sounds thus comprise nearly half of all consonant productions, underscoring their critical importance for accurate pronunciation.

5.3. Limitations

This study has limitations in its scope and methodology. First, its findings draw on a convenient and stratified sample of 25 second-year university full-time students whose major is English Language Studies, which limits their generalizability to many other types of Vietnamese learners of English. Methodologically, the use of a single examiner and limited speech materials meant not all phonetic contexts could be included. The scope was deliberately focused to enable detailed acoustic analysis. Despite these constraints, the study offers a precise, phonetically grounded reference that forms a valuable foundation for targeted teaching strategies.

References

[1] [Abker, I. A. A. (2020). Difficulties in Pronouncing English Morphemes among Saudi EFL Students at Albaha University. A Case Study in Almandag. *Arab World English Journal*, *11*(2), 395–410.

[2] Brown, H. D. (1994). *Principles of language learning and teaching* (Vol. 1). Prentice Hall.

[3] Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*, 161–70.

[4] Corder, S. P. (1971). IDIOSYNCRATIC DIALECTS AND ERROR ANALYSIS. *IRAL- International Review of Applied Linguistics in Language Teaching*, *9*(2), 147-160

[5] http://doi:10.1515/iral.1971.9.2.147

[6] Davenport, M., & Hannahs, S. J. (2020). *Introducing phonetics and phonology*. Routledge.

[7] Harmer, J. (2001). The practice of English language teaching. *London/New York*, *32*(1), 401-405.

[8] Hirch, R. (2016). ToPhonetics. *Pronunciation in Second Language Learning and Teaching Proceedings*.

[9] Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

[10] Kent, R. D., & Kim, Y. (2008). Acoustic analysis of speech. *The handbook of clinical linguistics*, 360–380.

[11] Labov, W. (1972). Some principles of linguistic methodology. *Language in society*, *1*(1), 97-120.

[12] Ladefoged, P., & Johnson, K. (2015). A Course in Phonetics (Seventh). *Cengage Learning*.

[13] Lam, T. T. K., & Thi, N. A. (2022). Common mistakes in pronouncing English consonant clusters: A case study of Vietnamese learners: A case study of Vietnamese learners. *CTU Journal of Innovation and Sustainable Development*, *14*(3), 32–39.

[14] Liang, S., & Fu, Y. (2020). Otter. ai. Los Altos, Otter. ai.

[15] Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press.

[16] Martimo, E. (2015). 'The Handbook of Language Variation and Change', JK Chambers and Natalie Schilling (eds.)(2013) Malden/Oxford: Wiley-Blackwell. Pp. 616. ISBN: 978-0-470-65994-6. *Sociolinguistic Studies*, *9*(1), 155-157.

[17] Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. *Language and speech*, *21*(3), 221–241.

[18] Nguyen, A. & Brouha, C. (1998). The production of word final consonants in English by L1 speakers of Vietnamese. *Working Papers in Linguistics*, *5*, 73–94.

[19] http://doi.org/lingclub/WP/currentXML.php?Vol=5

[20] Nguyen, H. D. (2019). Errors in word-final consonant pronunciation in Vietnamese English interlanguage. In *Proceedings of 15th International Conference on Humanities and Social Sciences*.

[21] Nguyen, N. (2008). Interlanguage phonology and the pronunciation of English final consonant clusters by native speakers of Vietnamese. *Unpublished master's thesis, Ohio University*.

[22] Styler, W. (2013). Using Praat for linguistic research. *University of Colorado at Boulder Phonetics Lab*.

[23] Richard, W. (1976). Predictors of pronunciation accuracy in second language learning. *26*(2), 233–253.

[24] Riaño, J. C. P. (2021). Phonetic and Phonological Transfer from Northern Vietnamese to English in Consonant Clusters and Voiceless Final Obstruents. *Epos: Revista de filología*, (37), 185–209.

[25] Roach, P. (2009). *English phonetics and phonology paperback with audio CDs (2): A practical course*. Cambridge university press.

[26] Sato, C. J. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning*, *34*(4), 43–58.

[27] Slowik, O., & Dung, D. T. H. (2022). Selected Pronunciation Issues of South Vietnamese English. *Open Journal of Modern Linguistics*, *12*(2), 226–237.

[28] Tim, F., & Paul, A. D. (2008). Solutions-Elementary Student's Book with MultiROM, 11–17.

[29] Tran, T. K. L., & Nguyen, A. T. (2022). Common mistakes in pronouncing English consonant clusters: A case study of Vietnamese learners. *Can Tho University Journal of Science*, *14*(3), 32–39

[30] Trudgill, P. (1974) *The Social Differentiation of English in Norwich*. Cambridge University Press.

[31] To, M. T. (2025a). *English morphology*. Ho Chi Minh City: Phu nu Viet Nam Press, Tri Tue Ltd.

[32] To, M. T. & Phan, V. Q. (2025b). *English phonetics and phonology*. Ho Chi Minh City: Thanh Nien Press, Tri Tue Ltd.

[33] Yule, G. (2020). *The study of language*. Cambridge University Press.