

English Pronunciation Challenges Among Female Santri of English Intensive Program at Pesantren Annur Probolinggo

Ali Wafa¹, Luluk Khoiriyah²

Universitas Islam Internasional Darullughah Wadda'wah¹

Akademi Manajemen Informatika Dan Komputer Taruna²

Correspondence Email : luluk.khoiriyah@amik-taruna.ac.id

Abstract

This descriptive qualitative study investigates the female santris' English segmental pronunciation errors, attributed to L1 phonological transfer from the dissimilar sound systems of English and Indonesian. Data from ten participants, collected via audio recordings, revealed systematic substitution and deletion processes affecting consonant and vowel phonemes. Consonantal errors included the substitution of /v/ with [f], /ð/ with [d] or [θ], /θ/ with [t] or [s], /tʃ/ with [c], /ʒ/ with [ʃ], /ʃ/ with [s], and /z/ with [s], alongside the deletion of /k/, /g/, /t/, and /s/ within consonant clusters. Vowel inaccuracies involved alterations in tongue height and length, such as raising /ɪ/ to /i/, shortening /i:/ to /e/ or /ɛ/, and elongating /o/ to /u:/, coupled with centralization and diphthongization processes. The findings indicate that these are not random errors but manifestations of a coherent, alternative phonological system. This system exhibits a preference for voiceless over voiced fricatives, stops over interdental fricatives, a marked simplification of complex consonant clusters, and a reorganized vowel space that avoids peripheral, tense, or distinct lax vowels. The study provides an empirical basis for targeted pedagogical interventions and contributes to establishing a detailed phonological error profile for Indonesian EFL learners in the *Pesantren* context.

Keywords: *consonant, English, female santri, pronunciation, vowel*

Abstrak

Studi kualitatif deskriptif ini menyelidiki kesalahan pelafalan segmental bahasa Inggris santri perempuan, diatribusikan melalui transfer fonologis L1 dari sistem bunyi bahasa Inggris dan Indonesia yang berbeda. Data diperoleh dari sepuluh partisipan, melalui rekaman audio, menunjukkan proses substitusi dan delesi sistematis mempengaruhi fonem konsonan dan vokal. Kesalahan konsonan mencakup substitusi /v/ dengan [f], /ð/ dengan [d] atau [θ], /θ/ dengan [t] atau [s], /tʃ/ dengan [c], /ʒ/ dengan [ʃ], /ʃ/ dengan [s], dan /z/ dengan [s], bersamaan dengan delesi /k/, /g/, /t/, dan /s/ dalam gugus konsonan. Ketidakakuratan vokal mencakup perubahan ketinggian dan panjang lidah, seperti menaikkan /ɪ/ menjadi /i/, pemendekan /i:/ menjadi /e/ atau /ɛ/, dan pemanjangan /o/ menjadi /u:/, yang disertai dengan proses sentralisasi dan diftongisasi. Temuan ini menunjukkan bahwa temuan ini bukanlah kesalahan acak, melainkan perwujudan dari sebuah sistem fonologis alternatif yang koheren. Sistem ini menandakan adanya preferensi terhadap frikatif nirsuara dibandingkan frikatif bersuara, hentian (stop) dibandingkan frikatif interdental, penyederhanaan yang nyata pada gugus konsonan kompleks, serta ruang vokal yang diatur ulang yang menghindari vokal perifer, tegang, atau vokal lunak yang berbeda. Studi ini memberikan dasar empiris untuk intervensi pedagogis yang terarah serta berkontribusi dalam menyusun profil kesalahan fonologis yang rinci bagi pelajar EFL Indonesia dalam konteks pesantren.

Kata kunci: *konsonan, Bahasa Inggris, santriwati, pelafalan, vokal*

Introduction

In Indonesia, English is acquired as a foreign language and is increasingly regarded as an essential competency for academic advancement, professional growth, and global engagement. Nevertheless, despite prolonged exposure through formal education, many learners encounter persistent difficulties in oral production, particularly in the domain of pronunciation. Pronunciation extends beyond the goal of native-like speech; it is a critical component of communicative efficacy, where inaccuracies can impede intelligibility, diminish speaker confidence, and compromise interactive success. From a theoretical perspective in English Language Teaching (ELT), this shift reflects the broader pedagogical transition from an accuracy-oriented, Audio-Lingual Method to reduce mispronunciation (Abrar & Ma'rifatulloh, 2025) towards Communicative Language Teaching (CLT) to motivate students to talk and actively participate in communication (Slobodiak, 2023). This paradigm change prioritizes intelligibility and comprehensibility as the primary goals of pronunciation instruction, emphasizing successful oral communication in authentic contexts over the attainment of native-like accuracy.

For Indonesian EFL learners, these difficulties primarily arise from phonological discrepancies between the English and Indonesian sound systems, a fundamental source of error as emphasized by Gilakjani (2011). English employs several phonemes absent in Indonesian, such as the dental fricatives /θ/ and /ð/ (as in *think* and *this*, respectively) and the palato-alveolar fricative /ʒ/ (as in *measure*). Consistent with Islam (2020) research, these non-native sounds are frequently substituted with the closest L1 phonological approximations. This phenomenon (English phonemes absent in Indonesian) can be explained by the theory of Cross-Linguistic Influence (CLI), which posits that a learner's first language (L1) phonological system systematically influences the acquisition of a second language (L2), leading to negative transfer or mispronunciation (Chen, 2022). Such influence is particularly evident in the production and perception of L2 phonetic segments and structures that are absent or manifest differently in the L1 inventory. Additional challenges include the production of consonant clusters, distinctions in vowel length, and mastery of lexical stress patterns, which adversely affect oral fluency (Chang, 2004; Gilakjani, 2011). Moreover, the influence of regional languages further complicates acquisition by affecting learners' phonological perception and production (Islam, 2020; Mathew, 1997).

The distinctive context of the English Intensive Program at Pesantren Annur Probolinggo, East Java, presents a valuable opportunity to explore these challenges. This immersive program within an Islamic boarding school enrolls students from diverse geographical and linguistic backgrounds with varied prior exposure to English. However, limited scholarly attention has been devoted to such intensive programs in Pesantren settings, which differ considerably from mainstream institutions. Consequently, the specific phonological features that pose the greatest difficulty here, along with tailored instructional approaches, remain underexplored. This study aims to fill this gap by investigating the English consonant and vowel phonemes frequently mispronounced by students in the English program at the Pesantren.

A principal difficulty for Indonesian learners involves mastering unfamiliar segmental features: the discrete vowels and consonants that form spoken language. The pedagogical significance of this line of inquiry stems from its potential to inform targeted instructional interventions. The identification of specific, recurrent segmental and suprasegmental errors constitutes a critical prerequisite for enhancing EFL learners' pronunciation intelligibility and overall communicative effectiveness. Accurate production is essential, as errors may lead to semantic shifts (Maraden & Silalahi, 2016). The English phonological system, with 44 phonemes, presents considerable perceptual and articulatory challenges due to its asymmetry with Indonesian. Among the most frequently problematic consonants include the dental fricatives /θ/ and /ð/, often substituted with /t/ and /d/ (Widya et al., 2025), voicing contrasts in word-final position (e.g., /p/ vs. /b/), which often lead to communication breakdowns (Yagi & Rosari, 2023), and the distinction between the labiodental fricatives /f/ and /v/, where /v/ is often replaced by /f/.

Vowel sounds also present considerable difficulty. Common points of confusion include the contrast between /ɪ/ and /i:/ (as in ship and sheep), /ʊ/ and /u:/ (as in full and fool), and the schwa /ə/, which is absent in Indonesian. (Nakjan et al., 2018) reported frequent confusion between /ɪ/ and /i:/ in intensive programs, while (Almutalabi, 2018) noted prevalent inaccuracies in diphthongs. This phenomenon of negative phonological transfer, where target sounds are substituted with L1 approximations, is well-established (Situmorang et al., 2023) and is exacerbated by the influence of regional dialects.

Effective instruction to address these issues should target both perception and production, using techniques like minimal pair tasks, articulatory training, and technology-enhanced tools like automatic speech recognition (ASR) for immediate feedback (Haghighi & Rahimy, 2017; Zuhri & Lizamuddin, 2025). Such pedagogical applications are underpinned by skill acquisition theory, which contends that declarative knowledge (i.e., phonetic rules) must be proceduralized through sustained practice and corrective feedback to attain automaticity in production (DeKeyser, 2014). Nevertheless, a notable research gap persists regarding investigating these phonological issues within intensive English programs in Pesantren settings like Pessantren Annur, particularly from the perspective of how specific ELT theories can inform and shape effective pronunciation pedagogy in this unique context.

Method

This study was designed to investigate pronunciation errors in English segmental features—specifically consonants and vowels—produced by ten participants from the Annur English Intensive Program during a speech contest. A descriptive qualitative methodology was employed to provide a detailed analysis of the phonological data, grounded in the disciplinary framework of phonetics. The primary instrument for data collection and analysis was the researcher, who conducted systematic observation and phonetic transcription of the participants' oral productions.

The data for this study were collected through direct observation and audio recording during a speech contest. The data collection procedure consisted of three systematic steps. First, the researcher attended the contest to observe live performances and record participants' speeches. Subsequently, the recordings were

meticulously transcribed, and all instances of mispronounced English phonemes were identified and annotated. Following the identification of mispronounced phonemes, these were transcribed utilizing the International Phonetic Alphabet (IPA). The transcriptions were subsequently compiled and subjected to an analytical procedure involving two independent raters to authenticate the precision of the identified phonological inaccuracies. Finally, the errors were categorized according to their segmental phonological features for further analysis.

The data analysis was conducted in two sequential stages. The first stage involved categorizing mispronounced English sounds according to their segmental phonological properties. All collected phonetic data were transcribed and systematically classified into two primary categories: consonant errors and vowel errors. The second stage entailed a detailed analysis of the identified pronunciation issues, focusing on the segmental characteristics of the mispronounced phonemes. To ensure analytical accuracy, reference was made to authoritative sources, including the Oxford Advanced Learner's Dictionary and established theories of English phonetics and phonology as articulated by scholars such as Roach, Jones, and Yule. Where pertinent, the phonological environments contributing to specific mispronunciation patterns were also examined to provide a more comprehensive account of the errors.

Result

To address the research question, the findings present an analysis of the data by categorizing pronunciation errors into two phonological classes: consonants and vowels. This classification elucidates the segmental features that constitute the primary pronunciation challenges encountered by the ten research participants during the speech contest.

The problem with the English consonants

The following consonant sounds were inaccurately produced by the participants during their research proposal presentations. Pronunciation accuracy was assessed by comparing their productions against the standard phonetic transcriptions provided in the Oxford Advanced Learner's Dictionary. The misarticulated consonants will be analyzed within the framework of key phonological theories proposed by (Jones & Jones, 1956; Roach, 1983; Yule, 1996) and Yule (1996), with specific attention to the dimensions of voicing, manner of articulation, and place of articulation.

a. The sound [v]

Table 1 presents a subset of lexical items in which the target phoneme /v/ was realized as the phoneme /f/ by the participant.

Table 1. Problem with the sound [f]

Position	Word	The correct pronunciation	The subject pronunciation
Initial	View	/vju:/	/fju:/
	Very	/'veri/	/'feri/
	Investigate	/ɪn'vestɪɡert/	/ɪn'fɛstɪɡert/
Medial	Seven	/'sɛv(ə)n/	/'sɛfən/
	Divide	/dɪ'vaɪd/	/dɪ'faɪd/

	Level	/'ləv(ə)l/	/'ləfəl/
	Convey	/kən'veɪ/	/kən'feɪ/
	Of	/ɒv/	/ɒf/
Final	Five	/(ə)v/	/faɪf/
	Solve	/sɒlv/	/sɒlf/

The English consonant /v/ is a voiced labiodental fricative. Data indicates that while participants generally used the correct labiodental placement and fricative manner, a prevalent error was the failure to maintain voicing. This devoicing resulted in the production of its voiceless counterpart, /f/. These two phonemes share an identical place and manner of articulation, differing only in voicing. This substitution error occurred in all three-word positions: initial, medial, and final. In the initial position, substitutions were found in words like “view” /vju:/ and “very” /'veri/, which were produced as /fju:/ and /'feri/. This error occurred when the target /v/ was immediately followed by a vowel, as seen in /'feri/, where /f/ is followed by /e/. A similar pattern was observed medially. For example, the /v/ in “seven” /'sev(ə)n/ and “level” /'ləv(ə)l/ was substituted, resulting in /'sefən/ and /'ləfəl/. This happened in various contexts, such as when /v/ was preceded by a consonant and followed by a vowel, as in pronouncing “investigate” as /ɪn'fɛstɪgeɪt/ (with /f/ preceded by /n/ and followed by /e/), or by a diphthong, as in “convey” becoming /kən'feɪ/. Finally, substitutions in the word-final position were observed in words like “five” /faɪv/ and “solve” /sɒlv/, which were produced as /faɪf/ and /sɒlf/. This devoicing occurred when the final [v] was preceded by a vowel, as in “of” becoming /ɒf/, or by a diphthong, as in /faɪf/.

b. The sound [ð]

Table 2 illustrates the substitution of the phoneme /ð/ by the participant, with realizations predominantly occurring as either /d/ or /θ/.

Table 2. Problem with the sound [ð]

Position	Word	The correct pronunciation	The subject pronunciation
Initial	The	/ðə/, /ði/, /ði:/	/də/
	Then	/ðen/	/den/
Final	with	/wɪð/	/wɪθ/

Analysis of participant errors revealed challenges with the English phoneme /ð/, a voiced dental fricative, in both word-initial and word-final positions. In the initial position, the primary error involved a deviation in the place and manner of articulation, while voicing was maintained. A notable substitution occurred where /ð/ was articulated as a voiced alveolar stop [d]. This was characterized by the tongue contacting the alveolar ridge rather than the teeth, altering the manner from fricative to plosive. This was systematically observed in words like “the” and “then,” realized as [də] and [den]. The data indicate this substitution occurred in a specific phonological context: when /ð/ was followed by a vowel. In final position, a distinct

error pattern emerged. The phoneme was produced with the correct manner and place—fricative and dental—but was consistently devoiced, resulting in a substitution with [θ]. This is exemplified in the word “with” /wɪð/, realized as /wɪθ/. This devoicing was conditioned by its context, occurring only when the sound was preceded by a vowel, as confirmed by the vowel [ɪ] immediately before the substituted fricative in “with.”.

c. The sound [θ]

The data presented in Table 3 illustrate common phonological substitutions made by the participants, specifically the replacement of the dental fricative /θ/ with the alveolar plosive /t/ in both word-initial and word-final positions, as well as the substitution of /θ/ with the alveolar fricative /s/ in word-initial position.

Table 3. Problem with the sound [θ]

Position	Word	The correct pronunciation	The subject pronunciation
Initial	Thank	/θaŋk/	/taŋk/
	Three	/θri:/	/tri:/
	Third	/θə:(r)d/	/sə:rd/
Final	Method	/'mɛθəd/	/'mɛtəd/

Phonetically, the sound /θ/ is characterized as a voiceless dental fricative. Accurate production of this phoneme requires the simultaneous fulfilment of three articulatory features: voicelessness, dental place of articulation, and fricative manner. However, certain subjects in the study deviated from two of these features when articulating /θ/. Specifically, the place of articulation was shifted from dental to alveolar. Moreover, the intended fricative manner was realized as a stop or plosive articulation. The voicing aspect, however, was produced correctly. These phonetic deviations resulted in the substitution of /θ/ with /t/, a voiceless alveolar stop. This substitution was observed in both initial and medial positions. For instance, in word-initial position, the target words “three” /θri:/ and “thank” /θaŋk/ were produced as /tri:/ and /tæŋk/, respectively. Analysis of the phonological environments revealed that the substitution of initial /θ/ by /t/ occurred under two distinct conditions: first, when followed by a consonant, as in /tri:/ where /t/ is followed by /r/; and second, when followed by a vowel, as in /tæŋk/ where /t/ is followed by /æ/.

d. The sound [tʃ]

Table 4 illustrates the target phoneme /tʃ/ substituted by the phoneme /c/ in medial position by the participant. Notably, such substitutions were not observed in either initial or final positions.

Table 4. Problem with the sound [tʃ]

Position	Word	The correct pronunciation	The subject pronunciation
Medial	Feature	/'fi:tʃə(r)/	/'fi:cə(r)/
	Switching	/swɪtʃɪŋ/	/swɪcɪŋ/

The phoneme /tʃ/ is a voiceless palato-alveolar affricate. Its accurate production depends on the coordinated execution of three articulatory features: voicing, place, and manner. The pronunciation difficulty observed among participants pertained specifically to the manner of articulation. Whereas /tʃ/ requires a complete obstruction followed by a gradual, fricated release, participants frequently

produced a stop consonant without the subsequent frication. This error occurred while both the voiceless quality and the palato-alveolar place were maintained. Consequently, the affricate /tʃ/ was substituted by the voiceless palatal stop [c], as these two sounds share identical places of articulation and voicing but differ in manner. This substitution was observed exclusively in medial positions. For instance, the /tʃ/ in words like “feature” (/ˈfi:tʃə(r)/) and “switching” (/ˈswɪtʃɪŋ/) was realized as [c], resulting in [ˈfi:cə(r)] and [ˈswɪcɪŋ]. A phonological analysis reveals this substitution occurs primarily in intervocalic contexts, where the target sound is flanked by vowels, as in the realization of “feature” as /ˈfi:cər/.

e. The sound [ʒ]

The following table illustrates the target phoneme [ʒ] substituted with [ʃ] by the participant, occurring exclusively in medial position.

Table 5. Problem with the sound [ʒ]

Position	Word	The correct pronunciation	The subject pronunciation
Medial	Conclusion	/kənˈklu:ʒ(ə)n/	/kənˈklu:ʃən/
	Cohesion	/kəʊˈhi:ʒ(ə)n/	/kəʊˈheʃən/

The consonant sound /ʒ/ is characterized by three distinct phonetic features: voicing, palatal articulation, and frication. The accurate production of /ʒ/ requires the simultaneous realization of all three features. Failure to maintain voicing results in the devoicing of /ʒ/, yielding the voiceless palato-alveolar fricative /ʃ/. This phonological error was observed among EFL participants specifically in medial position, affecting words such as “conclusion” /kənˈklu:ʒən/ and “cohesion” /kəʊˈhi:ʒən/, which were realized as /kənˈklu:ʃən/ and /kəʊˈheʃən/, respectively. It is noteworthy that vowel quality alterations also occurred in these productions; these will be addressed separately in the section devoted to vowel errors.

A phonological analysis of the data, as presented in the table, indicates that the substitution of the voiced post-alveolar fricative [ʒ] with its voiceless counterpart [ʃ] in medial position occurs within a specific phonological context. This segmental error is characterized by the target phoneme being intervocalic, meaning it is both preceded and followed by a vowel. This pattern is exemplified in the participant's production of the word “conclusion,” realized as /kənˈklu:ʃən/, wherein the erroneous [ʃ] is preceded by the vowel [u:] and followed by the schwa [ə].

f. The sound [ʃ]

Table 6 illustrates the substitution of the target phoneme /ʃ/ with /s/ by the participant, an error pattern observed in both word-initial and word-medial positions throughout the study.

Table 6. Problem with the sound [ʃ]

Position	Word	The correct pronunciation	The subject pronunciation
Initial	Show	/ʃəʊ/	/səʊ/
Final	Contribution	/kɒntriˈbju:ʃ(ə)n/	/kɒntriˈbu:sən/

The consonant [ʃ] is phonologically characterized as a voiceless palato-alveolar fricative. Its accurate production requires the simultaneous fulfillment of three articulatory features: voicelessness, palato-alveolar placement, and fricative manner. Several participants exhibited difficulty in the articulation of this sound, specifically

regarding its place of articulation. Rather than producing the sound at the palato-alveolar region, as required, participants frequently shifted the articulation to the alveolar ridge. Consequently, the output was realized as a voiceless alveolar fricative, phonetically equivalent to [s]. This substitution is phonologically plausible, as both [ʃ] and [s] share identical voicing and manner of articulation, differing primarily in their place of articulation.

The substitution of the phoneme /ʃ/ with /s/ was observed in both initial and medial word positions. A notable instance occurred in the production of the lexical item "show" /ʃəʊ/, where the target initial fricative /ʃ/ was replaced by /s/, resulting in the non-standard realization /səʊ/. This segmental error, specifically the devoicing and alveolarization of the post-alveolar fricative, occurred in a phonological environment preceding a diphthong. In this context, the substituted alveolar fricative /s/ was directly followed by the closing diphthong /əʊ/.

g. The sound [z]

Table 7 presents the target phoneme /z/ substituted by the phoneme /s/ in medial position, as produced by the participant during the data collection period.

Table 7. Problem with the sound [z]

Position	Word	The correct pronunciation	The subject pronunciation
Medial	Result	/rɪ'zʌlt/	/rɪ'sʌlt/
	Example	/ɪg'zɑ:mp(ə)l/	/ɛg'sɑ:mpəl/
	Present	/prɪ'zɛnt/	/prɪ'sɛnt/
	Examine	/ɪg'zæmɪn/	/ɛg'sæmɪn/

The phoneme /z/ is a voiced alveolar fricative consonant, the accurate production of which depends on the simultaneous articulation of three phonetic features: voicing, alveolar placement, and frication. Failure to adequately maintain any one of these articulatory components may result in the substitution or distortion of the target sound, leading to phonological inaccuracies typical among EFL learners. Several participants show difficulties in the production of the voiced alveolar fricative phoneme /z/. While they accurately replicated its manner and place of articulation, a consistent devoicing error occurred, resulting in the substitution of the voiceless alveolar fricative /s/. This phonological substitution was observed exclusively in medial position within words. For instance, the target words "result" (/rɪ'zʌlt/) and "present" (/prɪ'zɛnt/) were realized as /rɪ'sʌlt/ and /prɪ'sɛnt/, respectively. This error pattern highlights a specific challenge in maintaining voicing contrast among EFL learners, particularly affecting the accurate production of voiced obstruents in intervocallic contexts. Table 7 indicates that the substitution of the voiced alveolar fricative [z] with its voiceless counterpart [s] in medial position occurred across three distinct phonological contexts. Specifically, the devoicing of [z] to [s] was observed in intervocalic environments—that is, when preceded and followed by a vowel. For example, in the participant's production of the word "result" as /rɪ'sʌlt/, the target phoneme [z] was realized as [s], situated between the preceding vowel [ɪ] and the following vowel [ʌ].

h. The sound [k], [g], [t], and [s] deletion.

The sounds [k], [g], [t], and [s] exhibit distinct articulatory and phonological

characteristics. The voiceless velar plosive /k/ is produced by obstructing airflow at the velum without vocal fold vibration. Its voiced counterpart, /g/, shares this place and manner of articulation but involves vocal fold vibration. In contrast, /t/ is a voiceless alveolar plosive, articulated at the alveolar ridge. The sound /s/ is also voiceless and alveolar but is a fricative, produced with a narrow constriction that creates turbulent airflow. Consequently, the primary distinctions between these sounds lie in their voicing, place of articulation, and manner of production. Table 8 illustrates the specific deletion of the phoneme /k/ in post-consonantal positions.

Table 8. Problem with the sound [k]

Position	Word	The correct pronunciation	The subject pronunciation
Final	Task	/tɑːsk/	/tɑːs/

A phonological analysis of the data, as presented in Table X, indicates that the elision of the phoneme /k/ in the target word “task” occurred in a specific phonological environment. The deletion was consistently observed in a syllable-final position within a closed syllable, immediately following the fricative consonant /s/. This phenomenon resulted in the simplification of the /sk/ consonant cluster, yielding the phonetic realization [tɑːs] instead of the target form /tɑːsk/. The data suggest a phonological process wherein the velar plosive /k/ is susceptible to deletion when it is preceded by a consonant, specifically the alveolar fricative /s/, in a coda position. A parallel pattern of phonological simplification was observed with the phonemes /g/, /t/, and /s/. The research participants consistently modified consonant clusters in coda position within closed syllables by employing a deletion strategy, omitting these target phonemes when they occurred as the final consonant. Table 9 illustrates the deleted [g] sound following a preceding consonant in the speech output of EFL learners.

Table 9. Problem with the sound [g]

Position	Word	The correct pronunciation	The subject pronunciation
Medial	English	/'ɪŋɡlɪʃ/	/'ɪŋlɪʃ/
	Language	/'læŋɡwɪdʒ/	/'læŋwɪdʒ/

The table indicates that the deletion of the phoneme [g] occurs in syllable-final position within closed syllables and is consistently preceded by a consonant. For example, subjects produced the words “English” and “language” as [ˈɪŋlɪʃ] and [ˈlæŋwɪdʒ], respectively, omitting the [g] sound in both instances. This pattern suggests that the elision of [g] is conditioned by its phonological environment—specifically, when it appears medially between two consonants. Similarly, Table 10 further illustrates [t]-deletion following a consonant, indicating a broader phonological tendency for stop deletion in post-consonantal contexts.

Table 10. Problem with the sound [t]

Position	Word	The correct pronunciation	The subject pronunciation
Final	Text	/tɛkst/	/tɛks/
	Percent	/pə(r)'sɛnt/	/pər'sɛn/

The table indicates that the deletion of the final /t/ phoneme occurs in closed syllables when it is preceded by a consonant. For example, the subjects' realizations

of “percent” and “text” as /pər'sen/ and /teks/ demonstrate the elision of /t/ in post-consonantal position within a coda cluster. From these instances, a phonological pattern can be generalized: the deletion of word-final /t/ is conditioned by a preceding consonant. This is illustrated in the production of “percent” as /pər'sen/, where the omitted /t/ is immediately preceded by the nasal consonant /n/. Table 11 illustrates the deleted [s] sound following a preceding consonant in the speech output of EFL learners.

Table 11. Problem with the sound [s]

Position	Word	The correct pronunciation	The subject pronunciation
Final	Significance	/sɪg'nɪfɪk(ə)ns/	/sɪg'nɪfɪkən/

The data reveals that the elision of the alveolar fricative /s/ in the word “significance” occurred in a specific phonetic context: positioned in the coda of a closed syllable and following the alveolar nasal consonant /n/. This phenomenon represents a simplification of the consonant cluster /ns/ through the deletion of the final /s/, resulting in the surface form [sɪg'nɪfɪkən]. It should be noted that the concomitant vowel quality change from [ɪ] to [i] is not attributable to this particular deletion process. This pattern is consistent with a recognized phonological environment for consonant deletion, wherein the target segment is preceded by another consonant, the nasal /n/.

The problem with the English vowels

The misarticulations of vowel sounds will be analyzed within the theoretical frameworks of vowel phonetics as described by (Jones & Jones, 1956; Roach, 1983; Yule, 1996), and Yule (1996), with particular emphasis on tongue position and articulation.

a. The vowel [ɪ]

The vowel /ɪ/ is a close front vowel, articulated with the front of the tongue raised toward the hard palate and slightly spread lips. Its height is closer to a close-mid vowel, making it qualitatively lower and more centralized. Many participants struggled with this sound due to inaccurate tongue positioning, frequently leading to phonological substitution. Table 12 details these pronunciation errors for the vowel [ɪ].

Table 12. Problem with the sound [ɪ]

Position	Word	The correct pronunciation	The subject pronunciation
Initial	Examine	/ɪg'zæmɪn/	/ɛg'sæmɪn/
	Perfect	/'pə:(r)fɪkt/	/'pə:rfekt/
Medial	Preferred	/prɪ'fə:(r)d/	/prɛ'fə:rd/
	Examine	/ɪg'zæmɪn/	/ɪg'sæmɪn/
	This	/ðɪs/	/ðɪs/

The participants in the study exhibited two principal challenges in the production of the vowel sound [ɪ]. First, although they articulated the sound within the approximate front region of the vowel space, their tongues remained insufficiently lowered, resulting in the production of [i] rather than [ɪ]. This substitution error stemmed from the articulatory distinction between the two vowels: [i] is characterized by a higher tongue position compared to [ɪ]. Consequently, the

target phoneme [ɪ] was systematically replaced by [i]. For instance, in words such as “this” (/ðɪs/) and “examine” (/ɪgˈzæmɪn/), the medial [ɪ] was realized as [i], yielding erroneous pronunciations such as /ðis/ and /ɪgˈzæmɪn/.

b. The sound [i:]

The vowel [i:] is a long, close front vowel. Articulatorily, it requires the front of the tongue to be raised high towards the hard palate, with slightly spread lips. Its prolonged duration and higher tongue elevation distinguish it from its short counterpart, [ɪ]. This distinction is a common source of pronunciation difficulty for EFL learners. An analysis of participant production revealed errors in articulating [i:] in medial position. The data, summarized in Table 13, shows that a subset of learners exhibited two distinct types of mispronunciations of this target phoneme.

Table 13. Problem with the sound [i:]

Position	Word	The correct pronunciation	The subject pronunciation
	These	/ði:z/	/ðez/
Medial	Cohesion	/kəʊˈhi:ʒn/	/kəʊˈhɛ:sn/
	Thesis	/ˈθi:sis/	/ˈtɛ:sis/

The target phoneme /i:/ was frequently substituted with the vowels /e/ and /ɛ/. The primary substitution, resulting in the production of /e/, was attributed to a misarticulation involving tongue height and vowel length. Specifically, the front of the tongue was lowered from the high position required for the close front vowel /i:/ to a mid-high position characteristic of a close-mid vowel. Concurrently, the vowel duration was shortened. This articulatory error is exemplified in the word “these” /ði:z/, where the substitution of the nucleus /i:/ with /e/ yielded the erroneous pronunciation /ðez/.

c. The sound [ɛ]

The vowel /ɛ/ is a short, open-mid front vowel, articulated with the front tongue in a mid-position and slight lip spreading. Participants demonstrated difficulty accurately producing this sound. Analysis identified two distinct error patterns in its realization. The specific phonological deviations observed during the study are presented in Table 14.

Table 14. Problem with the sound [ɛ]

Position	Word	The correct pronunciation	The subject pronunciation
	general	/ˈdʒɛnərəl/	/ˈdʒənəɾəl/
Medial	generally	/ˈdʒɛnərəli/	/ˈdʒənəɾəli/
	Said	/ˈsed/	/ˈseɪd/

First, the target phoneme /ɛ/, a front vowel, was frequently realized by participants using a central tongue articulation. Specifically, the tongue body was elevated to a mid-central position, resulting in the production of the central vowel schwa /ə/. This phonological substitution occurred in the initial syllables of the lexical items general /ˈdʒɛnərəl/ and generally /ˈdʒɛnərəli/, which were consequently produced as [ˈdʒə.nə.rəl] and [ˈdʒə.nə.rə.li]. This error pattern also involved epenthesis, the insertion of an additional vowel sound, a phenomenon which will be examined in the subsequent section.

d. The sound [ʊ]

The vowel /ʊ/ is a near-close near-back sound. It is articulated with the back of the tongue raised and retracted toward the soft palate, but not to the full extent of a close vowel. Its production also features slight lip rounding. Common mispronunciations of this phoneme, as detailed in Table 15, were observed among the study's participants.

Table 15. Problem with the sound [ʊ]

Position	Word	The correct pronunciation	The subject pronunciation
	Look	/lɒk/	/lu:k/
Medial	Good	/gʊd/	/gu:d/
	Book	/bʊk/	/bu:k/

Several participants exhibited a common phonological error in their production of the vowel /ʊ/. The mispronunciation was characterized by a retraction and excessive elevation of the dorsal region of the tongue, coupled with moderate lip rounding. This articulatory configuration resulted in the substitution of the target /ʊ/ with the vowel /u:/, thereby elongating the vowel duration. This error pattern was observed specifically in medial word positions. Consequently, lexical items such as "look" /lɒk/, "good" /gʊd/, and "book" /bʊk/ were realized as /lu:k/, /gu:d/, and /bu:k/, respectively.

e. The sound [ʌ]

The vowel /ʌ/ is an open-mid central unrounded sound, articulated with the tongue's central part slightly raised. This specific configuration is often challenging for EFL learners, who tend to substitute it with a more open or back vowel. Table 16 presents the phonological errors in /ʌ/ production identified in this study.

Table 16. Problem with the sound [ʌ]

Position	Word	The correct pronunciation	The subject pronunciation
	Multiple	/'mʌltɪpl/	/'maltɪpl/
Medial	Public	/'pʌblɪk/	/'pabɪk/
	Construct	/kən'strʌkt/	/kən'strakt/

Several participants exhibited a specific phonological error in their production of the vowel sound /ʌ/. The error was characterized by an excessive lowering of the tongue, resulting in a fully open vocal tract. This articulation deviates from the target pronunciation, which requires a slightly raised tongue position in the central region of the oral cavity, consistent with an open-mid vowel quality. Consequently, the target phoneme /ʌ/ was systematically substituted with the open central vowel /a/. This substitution error was observed in the medial position of the words multiple (/'mʌltɪpl/ → ['maltɪpl]), public (/'pʌblɪk/ → ['pabɪk]), and construct (/kən'strʌkt/ → [kən'strakt]).

f. The sound [ɜ:]

The vowel sound [ɜ:] is a long, mid-central vowel, characterized by its sustained duration relative to short vowels. Articulatorily, it is produced with the central part of the tongue raised to a position intermediate between open and close vowel heights, specifically approximating the open-mid region. The lip configuration remains neutral throughout its production. Table 17 illustrates the modified [ɜ:] realizations observed among the participants during the study.

Table 17. Problem with the sound [ɜ:]

Position	Word	The correct pronunciation	The subject pronunciation
	Heard	/hɜ:(r)d/	/hi:ɜrd/
Medial	Word	/'wɜ:(r)d/	/'wɒ:rd/
	World	/'wɜ:(r)ld/	/'wɒ:rld/

The subject exhibited a phonological deviation characterized by the diphthongization and reduction of the vowel length in /ɜ:/. Specifically, an epenthetic [i:] was inserted preceding the target vowel, resulting in a truncated and diphthongized realization approximating [i:ɜ]. This phonetic alteration was notably observed in the production of the word heard /hɜ:d/, which was articulated as [hi:ɜrd], indicating a substitution of the monophthong /ɜ:/ with the diphthongal variant [i:ɜ].

Discussion

The pronunciation challenges observed among the research participants primarily involved the substitution of English consonants and vowels with phonetically similar alternatives from their first language (L1) or sounds that aligned with a word's spelling. For consonants, the affected sounds included [v], [ð], [θ], [tʃ], [ʒ], [ʃ], and [z]. These substitutions arose from modifications in phonetic features like voicing, manner, or place of articulation. A common pattern was devoicing, where voiced consonants were replaced by their voiceless counterparts. For instance, participants frequently pronounced /'veri/ as /'fəri/ and /kən'klu:ʒən/ as /kən'klu:ʃən/, substituting [v] with [f] and [ʒ] with [ʃ]. These findings align with previous studies by (Rahman & Tralala, 2021; Rosyidah, 2014) and (Rosyidah, 2014). Participants also exhibited challenges with vowel and diphthong production. Common substitution errors occurred with vowels such as /i:/, /ɪ/, /e/, /ʊ/, /ʌ/, /ɜ:/, /ɒ/, /ɔ:/, and /ə/, consistent with Rahman & Tralala (2021) finding. For example, the word “look” /lʊk/ was often produced as /lu:k/, and “public” /'pʌblɪk/ as /'pabli:k/, replacing [ʊ] with [u:] and [ʌ] with [a]. A strong tendency toward orthographic interference was also evident. Learners frequently substituted vowel sounds with others that matched the written form. This was particularly noticeable with the schwa /ə/, which was often realized as a full vowel. For instance, “isolation” (/ˌaɪsəˈleɪʃən/) was pronounced as /ˌaɪsəˈleɪʃən/, influenced by the letter "o", and “relevant” (/ˈrɛləvənt/) as /ˈrɛləvənt/, following the spelling with "a". In summary, the segmental errors were predominantly characterized by sound substitution. The most common patterns involved replacing target English phonemes with acoustically similar L1 sounds or with vowels that correspond to orthographic representations, which may significantly impact intelligibility.

Conclusion

This study shows that English Intensive students at Pesantren Annur exhibited pronunciation errors involving several consonants, vowels, and diphthongs. For instance, two primary types of consonant-related pronunciation issues: the substitution of target consonants with phonetically similar sounds, and the deletion of consonants within clusters. Also, Specific substitution errors included the replacement of [v] with [f], [ð] with [d] or [θ], [θ] with [t] or [s], [tʃ] with [c], [ʒ]

with [ʃ], [ʒ] with [s], and [z] with [s]. Regarding deletions, the consonant sounds [k], [g], [t], and [s] were frequently omitted when occurring in consonant clusters. In addition, the common vowel errors among EFL learners involve inaccurate tongue positioning and substitutions, such as raising /ɪ/ to /i/, shortening /i:/ to /e/ or /ɛ/, and elongating /o/ to /u:/. Centralization occurs (e.g., /ɛ/ to /ə/), and /ʌ/ is often lowered to /a/, while /ɜ:/ may break into diphthongs like [i:ɜ] or shift toward /ɒ/.

This study holds significant implications for educators and curriculum designers, including those within English Intensive programs in Islamic boarding schools (*pesantren*), as it enables the prioritization of instruction on sounds empirically identified as most problematic for learners. Furthermore, this research contributes to the establishment of a specific phonological error profile for Indonesian learners, particularly within the *pesantren* context—a socio-linguistic environment characterized by its remarkable diversity and richness. Future research should investigate how the unique linguistic environment of the *pesantren*, including its daily routines and focus on Arabic, facilitates or hinders the acquisition of English phonology.

References

- Abrar, M. F., & Ma'rifatulloh, S. (2025). The Effectiveness of Audio- Lingual Method on Students Pronunciation Performance. *CENDEKIA: Jurnal Ilmu Sosial, Bahasa Dan Pendidikan*, 5(1), 263–272. <https://doi.org/10.55606/CENDEKIA.V5I1.3525>
- Almutalabi, M. M. (2018). Analysing the Improper Pronunciation of Diphthongs by Iraqi EFL learners. *JURNAL ARBITRER*, 5(1), 17–22. <https://doi.org/10.25077/AR.5.1.17-22.2018>
- Chang, F.-C. (2004). Chinese-speaking EFL learners' performances of processing English consonant clusters. *Proceedings of International Conference on English Instruction and Assessment*.
- Chen, Y. (2022). The Role of Cross-linguistic Influence in Second Language Acquisition. *Journal of Higher Education Research*, 3(1), 98. <https://doi.org/10.32629/JHER.V3I1.659>
- DeKeyser, R. (2014). Skill acquisition theory. In *Theories in Second Language Acquisition: An Introduction*. <https://doi.org/10.4324/9780203628942-11>
- Gilakjani, P. (2011). A Study on the Situation of Pronunciation Instruction in ESL/EFL Classrooms. *Journal of Studies in Education*, 1(1), 4. <https://doi.org/10.5296/JSE.V1I1.924>
- Haghighi, M., & Rahimy, R. (2017). The Effect of L2 Minimal Pairs Practice on Iranian Intermediate EFL Learners' Pronunciation Accuracy. *International Journal of Research in English Education*, 2(1). <https://doi.org/10.18869/acadpub.ijree.2.1.42>
- Islam, S. M. (2020). SEGMENTAL ERRORS IN ENGLISH PRONUNCIATION OF NONNATIVE ENGLISH SPEAKERS. *Journal of Education and Social Sciences*, 16(1), 14–24.
- Jones, L. G., & Jones, D. (1956). An Outline of English Phonetics. *Language*, 32(3). <https://doi.org/10.2307/410583>
- Maraden, R., & Silalahi, P. (2016). Pronunciation Problems of Indonesian EFL Learners: An Error Analysis. In *Journal of English Language and Culture* (Vol. 6, Issue 2).
- Mathew, I. B. (1997). Errors in pronunciation of consonants by Indonesian, Gayo and Acehnese learners of English as a foreign language. *Theses: Doctorates and Masters*. <https://ro.ecu.edu.au/theses/904>
- Nakjan, S., Syamsuar, F., Yenprem, W., Nakjan, M., & Pankaew, P. (2018). A Comparative Study of Strategy to Improve Problematic English Pronunciation Produced by Thai and Indonesian Young Learners: A Comparative Study of Strategy to Improve Problematic English Pronunciation Produced by Thai and Indonesian Young Learners. *Phetchaburi Rajabhat University*, 12(12). <https://article.pbru.ac.th/index.php/ENG/article/view/144>
- Rahman, A., & Tralala, L. N. (2021). EFL Students' Pronunciation Problems in Presenting Thesis Proposal at Tertiary Level of English Department. *Jurnal Kependidikan: Jurnal Hasil*

- Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, 7(3). <https://doi.org/10.33394/jk.v7i3.3921>
- Roach, P. (1983). *English Phonetics and Phonology A practical course Fourth edition*. www.cambridge.org
- Rosyidah, R. U. (2014). *An analysis of Phonetic Interferences on students' English Pronunciation in Shobahul Lughoh Program of Ma'had Sunan Ampel Al Aly*.
- Situmorang, J. J., Lubis, Y., & Lubis, R. F. Y. (2023). Phonological Challenges and Error Patterns in English Consonant Production by Indonesian Learners: A Literature Review. *Jurnal Pendidikan Dan Sastra Inggris*, 3(2). <https://doi.org/10.55606/jupensi.v3i2.1986>
- Slobodiak, S. (2023). On the communicative approach to the English language teaching in the context of blended learning. *Scientific Bulletin of the Izmail State University of Humanities. Section «Pedagogical Sciences»*, (62). [https://doi.org/10.31909/26168812.2023-\(62\)-25](https://doi.org/10.31909/26168812.2023-(62)-25)
- Widya, S., Sabbu, S., Rahman, N. S., Thalib Mbau, A., & Nurlaela, N. (2025). Pronunciation of English Fricatives by Indonesian EFL Learners: Patterns and Pedagogical Implications. *English Education and Literature Journal*, 5(02), 310–318. <https://doi.org/10.53863/E-JOU.V5I02.1749>
- Yagi, J. I., & Rosari, M. D. (2023). English Consonant Mispronunciation by Jakarta International University's Final Year English Department Students. *EDULINK EDUCATION AND LINGUISTICS KNOWLEDGE JOURNAL*, 5(1). <https://doi.org/10.32503/edulink.v5i1.3441>
- Yule, G. (1996). *The Study of Language*. Cambridge, 3(September).
- Zuhri, M. N., & Lizamuddin, A. (2025). The Influence of Speech Recognition Software on English Pronunciation Improvement: A Quasi-Experimental Investigation. *Jurnal Multidisiplin Dehasen (MUDE)*, 4(3), 549-556-549–556. <https://doi.org/10.37676/MUDE.V4I3.8409>