

PHONOLOGICAL ERROR ANALYSIS OF EFL STUDENTS' PRONUNCIATION IN ENGLISH SONGS USING SPEECH-TO-TEXT TECHNOLOGY TOOLS

Naliza Aiya^{1*}, Fitrah Syakila Rambe², Zulio Diva Ananta³, Yani Lubis⁴

^{1,2,3,4}Universitas Islam Negeri Sumatera Utara

nalizaaiya00@gmail.com

ABSTRACT

This research is motivated by the importance of mastering phonology in learning English as a foreign language, especially in the aspect of pronunciation which is still a challenge for students. This study aims to identify the types of phonological errors that appear in students' singing, determine the most dominant categories of errors, as well as uncover the factors that affect their pronunciation difficulties. The research uses a qualitative descriptive approach with an error analysis method. Data was obtained from the singing recordings of 19 students of the English Education Study Program which were then transcribed using several automatic transcription software and verified through phonetic analysis based on the International Phonetic Alphabet. The results showed that of the 19 identified deviations, there were 12 cases that could be categorized as representative phonological errors. These errors include consonant errors, vowel errors, sound omissions, sound additions, metathesis, epenthesis, and nasal errors. The most dominant categories of errors are consonant errors and vowel errors, which each account for 25% of the overall data. The findings also show that students' pronunciation difficulties are influenced by first language interference, limitations of phonological awareness, difficulty processing interconnected speech, and the influence of auditory imitation when singing English songs. This study concludes that phonological analysis combined with automatic transcription technology is able to provide a more objective picture of students' pronunciation error patterns and can be used as a basis for the development of more effective pronunciation learning.

Keywords: phonological errors, English pronunciation, EFL students, automatic transcription, phonetic analysis

INTRODUCTION

Learning English as a Foreign Language (EFL) today requires a mastery of communicative phonology. Phonological competence includes the ability to accurately produce segmental and suprasegmental features to achieve intelligible speech in global interactions. Fundamentally, learning a second language is not a linear or simple process, but rather a long and complex endeavor involving cognitive, affective, and social dimensions. Brown (2007, p. 15) has emphasized that language learning is not just a mastery of mechanical steps, but an effort that requires the total involvement of all aspects of the learner's self in order to be able to communicate spontaneously and meaningfully.

In the microlinguistic spectrum, Roach (2009, p. 31) explains that phonemes are the basic units that distinguish meaning, so an accurate understanding of phoneme systems and their transcription is a crucial foundation. However, for EFL learners, challenges often arise due to inaccuracies in articulation. (Koshy Joseph, 2023) has affirmed that effective communication depends not only on vocabulary and grammar, but also on pronunciation accuracy. Common mistakes, such as vocal distortion, pronunciation of silent letters, as well as stress and intonation patterns, are often obstacles that lower student confidence. This problem is in line with the findings of Salim (2025, p. 2), which shows that students often face difficulties in distinguishing certain consonant sounds, especially in words that contain silent consonant sounds. Furthermore, Gerosa et al. (2009, p. 1) noted that acoustic variability due to differences in vocal tract anatomy and imperfect articulator control is often a technical obstacle in learners' speech recognition.

Gaps in attention to phonological aspects risk hindering the effectiveness of long-term communication. Kelly (2000, p. 13) highlights a paradox in which phonology teaching is often neglected or is only done reactively. To achieve comprehensive mastery, phonology features

must be strategically integrated into learning planning so that they are not just complementary activities (Kelly, 2000, p. 14). At Swadhipa Natar High School, the low mastery of phonology is exacerbated by the reliance on less attractive textbook media, which contributes to the low achievement of students' speaking scores—with 61.08% of students not achieving the minimum completeness criteria (Salim, 2025, p. 4).

To overcome these challenges, the integration of Automatic Speech Recognition (ASR) technology offers an objective diagnostic solution. ASR is the process of converting acoustic signals captured through a microphone into a set of words that can be interpreted textually (Das et al., 2015, p. 191). This system operates through several crucial stages, ranging from acoustic signal processing, feature extraction using Mel Frequency Cepstral Coefficients (MFCC), to modeling using the Hidden Markov Model (HMM) to map sounds into word units (Das et al., 2015, p. 194). The main advantage of this technology-based approach is its ability to map phonological deviations consistently, which is very useful for educators to provide measurable feedback (Das et al., 2015, p. 192).

The integration of this technology is strengthened by the use of English songs as a learning medium. Shen (2009, p. 89) states that songs create a relaxed learning environment and activate both hemispheres of the brain. Songs provide rich exposure to connected speech, as well as lower students' affective filters. Misa (2024, p. 151) confirms that this technique is effective in improving pronunciation skills and learning motivation. This superiority is also attested by Salim (2025, p. 3), who found that songs function as a medium that trains students' hearing and articulation in a fun way due to the natural repetition of lyrics.

Although the use of songs in learning English pronunciation has been extensively researched, most previous research has focused on improving pronunciation skills, learning motivation, or the effectiveness of songs as a learning medium. Research that specifically identifies the types of phonological errors of EFL learners through the integration of speech-to-text technology and phonetic analysis based on the International Phonetic Alphabet (IPA) is still relatively limited. Therefore, this study seeks to fill this gap by presenting a more objective and measurable diagnostic approach. The novelty of this research lies in three main aspects. First, this study utilizes several automated transcription tools, namely TurboScribe AI, Instant Transcription, and Audio-to-Text Transcription, to obtain more comprehensive error identification results. Second, this study combines speech-to-text technology with IPA-based phonetic analysis so that pronunciation errors are not only detected automatically, but can also be verified linguistically. Third, this study uses English songs as an analytical context, thus allowing the observation of sound production in a more authentic situation and close to language learning practices.

This research aims to explore students' phonological errors through the integration of automatic transcription technology and analysis based on the International Phonetic Alphabet (IPA). The focus of the main problem is how speech-to-text technology can objectively map phonological deviations when students sing English songs, with a special emphasis on the silent consonant patterns that are often an obstacle for learners (Salim, 2025, p. 5). The use of optimized ASR technology has been proven to improve diagnostic precision in detecting pronunciation deviations compared to generically designed systems (Das et al., 2015, p. 192).

Based on this background, this research is directed to answer three main questions, namely: (1) what are the types of phonological errors that appear in student singing; (2) which phonological errors are the most dominant; and (3) what factors are behind the difficulties. This research is expected to make a methodological contribution for educators in accurately mapping phonological errors, as well as a guide for the development of more measurable song-based teaching materials.

METHODS

This study uses a qualitative descriptive approach, where according to Moleong (2005, p. 48), this approach is a research procedure that produces descriptive data in the form of written or spoken words from people and observed behaviors to understand the phenomenon in depth. In line with this, this study applies the Error Analysis method to classify the types of language errors that occur in the language production process by learners (Al-Sowaidi, 2022).

Participants in this study were 19 students in the fourth semester of the English Education Study Program of UIN North Sumatra who were chosen because they were assumed to have taken the Phonology or Pronunciation course. The primary data sources of the study consisted of audio recordings of English song singing produced by students as well as the original lyrics of songs used as a comparative reference in the process of identifying phonological errors.

The data collection technique is carried out through the documentation method by asking students to record English songs independently and then analyze. In this study, the researcher acted as the main instrument (human instrument) supported by several speech-to-text software, namely TurboScribe AI, Instant Transcription, and Transkriptor. The use of more than one software aims to improve the validity of the data through triangulation of transcription results so that error identification does not depend solely on a single speech recognition system. In addition, the use of multiple tools allows researchers to compare the transcription results produced by each system to obtain a more comprehensive picture of phonological errors.

The data analysis technique follows the interactive analysis model of Miles and Huberman (1992, p. 15) which includes data collection, data reduction, data presentation, and conclusion drawn. The research procedure began with the collection of audio recordings of student songs. Furthermore, the recordings are processed using TurboScribe AI, Instant Transcription, **and Transkriptor** to generate automatic transcriptions. The transcription results are then compared with the original lyrics of the song to identify words that have deviations or inconsistencies. The next stage is an initial screening using software to detect possible pronunciation errors (Al-Sowaidi, 2022), which is then followed by manual verification through rehearing of the recording to ensure data accuracy. After that, the errors found are transcribed into the International Phonetic Alphabet (IPA) to analyze the form of phonological deviation in more detail. In the final stage, errors are classified by their type, analyzed to find dominant patterns, and interpreted to identify the factors behind the occurrence of the error. Through this stage, the research seeks to distinguish between systematic errors and unintentional errors (Al-Sowaidi, 2022), so that a more accurate picture is obtained of the characteristics of students' phonological errors in singing English songs.

RESULTS AND DISCUSSION

Overview of Identified Errors

Tabel 1. Phonological Errors Identified from Students' English Song Performances

Subject	Song Title	Error Output	Target Form	IPA Diagnosis
1	<i>Blank Space</i>	"Where you been"	"Where have you"	Omission: Deletion of auxiliary /həv/.

2	<i>10,000 Hours</i>	"Grow up born"	"Grow up on "	Consonant Addition: Intrusive bilabial plosive /b/.
3	<i>Blue</i>	"Wait you by"	"Watch you go by"	Phonemic Substitution: /wɒtʃ/ replaced by /weɪt/.
4	<i>Payphone (versi no rep)</i>	"Bridges burned down"	"Bridges burn "	Morpho-phonemic Error: Overgeneralization of /-ed/ suffix.
5	<i>Perfect</i>	"Paid in what I see"	" Faith in what I see"	Devoicing: Labiodental fricative /f/ replaced by plosive /p/.
6	<i>Simple Plan Perfect</i>	"I love No"	"I know "	Phonetic Interference: Confusion between /noʊ/ and /lʌv/.
7	<i>Just a Friend to you</i>	"Very captured"	" Burying "	Labiodental Confusion: /b/ (bilabial) replaced by /v/.
8	<i>I'd Risk It All for You</i>	"Wind thing"	" One thing"	Vowel Shift: /ʌ/ (central) shifted to /ɪ/ (front).
9	<i>A Thousand Years</i>	"Time has broke"	"Time has brought "	Monophthongization: Diphthong /oʊ/ used for long vowel /ɔ:/.
10	<i>Can't Take My Eyes Off You</i>	"Blood on the way"	" Love in the way"	Cluster Error: Poor articulation of /lʌv ɪn/ sequence.
11	<i>The One That Got Away</i>	"RICO soon"	" Read you soon"	Assimilation Error: Slurred /d/ + /j/ sounds.
12	<i>The One That Got Away</i>	"1 day, go away"	"One that got away"	Syllable Deletion: Removal of functional segments.

13	<i>The Wheels on the Bus</i>	"All the 2, the town"	"All through the town"	Dental Fricative Error: /θ/ replaced by plosive /t/.
14	<i>That somebody</i>	"Expect someone"	" Except someone"	Metathesis: Transposition of /p/ and /s/ sounds.
15	<i>Remember Me</i>	"Though I have to"	"Though"	Epenthesis: Unnecessary sound/word addition.
16	<i>You'll Be in My Heart</i>	"Mean my heart"	" In my heart"	Nasal Nasalization: Alveolar /n/ shifted to bilabial /m/.
17	<i>Shape of My Heart</i>	"Put out the globe"	"Put out the glow "	Final Consonant Addition: Intrusive /b/ at word-end.
18	<i>Twinkle Twinkle Little Star</i>	"Above world"	"Above the world"	Article Omission: Deletion of the determiner /ðə/.
19	<i>Ours to Keep</i>	"Son of the day"	" Sun of the day"	Vowel Merger: Confusion between /ʌ/ and /ʊ/.

(Source: EFL Student Voice)

Table 1 presents all the phonological aberrations identified in the recording. However, not every deviation is chosen to be discussed in detail. Following the principle of qualitative error analysis, only representative cases that clearly show the phonological process are selected for in-depth analysis. Some cases are excluded from detailed discussion because they may reflect lexical substitution, semantic reinterpretation, or possible inaccuracies generated by speech-to-text converter applications, rather than actual phonological errors. Therefore, the next discussion focuses on the most linguistically representative examples to ensure analytical accuracy and theoretical consistency (Ellis & Barkhuizen, 2005, p. 57).

B. Representative Phonological Errors

Of the total 19 error data found in Table 1, only 12 were selected as representative data for further phonological analysis. This selection was made because not all errors that appear can be categorized as pure phonological errors. Some data involve changes in grammatical structure, omissions of functional words, or possible results of hearing errors that are not directly related to the production of language sounds. Therefore, this study focuses on data that show real sound changes, such as consonant substitution, vowel changes, sound addition, sound removal, metathesis, and epenthesis. These categories correspond to the phonological processes commonly discussed in segmental phonology studies (Odden, 2013, pp. 131–145).

Tabel 2. Pure Phonological Errors

Subject	Song Title	Error Output	Target Form	IPA Diagnosis
2	<i>10,000 Hours</i>	"Grow up born"	"Grow up on"	Consonant Addition: Intrusive bilabial plosive /b/.
5	<i>Perfect</i>	"Paid in what I see"	" Faith in what I see"	Devoicing: Labiodental fricative /f/ replaced by plosive /p/.
8	<i>I'd Risk It All for You</i>	"Wind thing"	" One thing"	Vowel Shift: /ʌ/ (central) shifted to /ɪ/ (front).
9	<i>A Thousand Years</i>	"Time has broke"	"Time has brought "	Monophthongization: Diphthong /oʊ/ used for long vowel /ɔ:/.
13	<i>The Wheels on the Bus</i>	"All the 2, the town"	"All through the town"	Dental Fricative Error: /θ/ replaced by plosive /t/.
14	<i>That somebody</i>	"Expect someone"	" Except someone"	Metathesis: Transposition of /p/ and /s/ sounds.
16	<i>You'll Be in My Heart</i>	"Mean my heart"	" In my heart"	Nasal Nasalization: Alveolar /n/ shifted to bilabial /m/.
17	<i>Shape of My Heart</i>	"Put out the globe"	"Put out the glow "	Final Consonant Addition: Intrusive /b/ at word-end.
19	<i>Ours to Keep</i>	"Son of the day"	" Sun of the day"	Vowel Merger: Confusion between /ʌ/ and /ɒ/.

Total = 9 cases

Tabel 3. Borderline Phonological Errors

Subject	Song Title	Error Output	Target Form	IPA Diagnosis
1	<i>Blank Space</i>	"Where you been"	"Where have you"	Omission: Deletion of auxiliary /həv/.
15	<i>Remember Me</i>	"Though I have to"	"Though"	Epenthesis: Unnecessary sound/word addition.
18	<i>Twinkle Twinkle Little Star</i>	"Above world"	"Above the world"	Article Omission: Deletion of the determiner /ðə/.

Total = 3 cases

Types of Phonological Errors That Appear in Student Singing

Based on the analysis of 12 representative data in Table 2, seven types of phonological errors were found that appeared in student singing, namely consonant errors, vowel errors, omissions, additions, metathesis, epenthesis, and nasal errors.

Consonant Errors

Consonant errors are one of the most common types of errors found in data. These errors include substitutions, additions, or changes to certain consonant characteristics.

For example, in faith data pronounced paid, the voiceless labiodental fricative phoneme /f/ is replaced by the voiceless bilabial explosive phoneme /p/. This error indicates a process of consonant substitution that is common in English learners as a foreign language when the target sound is considered difficult to produce (Avery & Ehrlich, 1992, p. 15).

Another example is found in the pronunciation of *through* to *two*, where the dental fricative phoneme /θ/ is replaced by the alveolar explosive phoneme /t/. The /θ/ sound is not present in the Indonesian phonological system so it is often replaced by the sound that is considered closest by Indonesian speakers (Odlin, 1989, p. 27).

In addition, there were also cases of adding consonants to the phrase *grow up on* which were pronounced to *grow up born* and *glow* which was pronounced as *globe*. In both cases, the phoneme /b/ appears that is not present in the target form.

Vowel Errors

The next type of error is vowel error which includes vowel shift, vowel fusion, and monophthongization. In the data of *one thing* pronounced as *wind thing*, the vowel /ʌ/ shifts to /ɪ/. This phenomenon indicates a vowel shift, which is a change in the position of the vocal articulation from the central position to the front position.

Meanwhile, in the data of *sun* pronounced as *son*, there is a vowel merger between the vowels /ʌ/ and /ɒ/. This mistake shows that students have difficulty distinguishing two English vowels that do not have a clear equivalent in Indonesian.

Another case was found in the word *brought* which is pronounced broke. In this data, there is a simplification of diphthongs to monophthongization, which is when complex vowel circuits are replaced with single vowels that are easier to produce (Odden, 2013, p. 143).

Omission

Omission is the omission of sounds or speech elements that should appear in the form of a target. In this study, omissions were found in the data *where have you* which was pronounced as *where you* and *above the world* which was pronounced as *above world*.

According to Yule (2020, p. 43), the omission of speech elements often occurs when the speaker processes speech so quickly that some phonological elements are not perceived in their entirety.

Addition

Addition occurs when the speaker adds a sound that is not present in the target shape. An example is found in the phrase *grow up on* which is pronounced to *grow up born*. The addition of the phoneme /b/ results in a new form that is different from the original target.

According to Fromkin, Rodman, and Hyams (2018, p. 287), the process of adding sounds can appear as an attempt by the speaker to adjust sound patterns that are considered easier or more familiar.

Metathesis

Metathesis is the exchange of positions of two sounds in a word. In the except data that is pronounced as *expect*, there is a shift in the position of the /s/ and /p/ sounds.

Metathesis is one of the phonological processes that occurs when the speaker has difficulty maintaining the correct phoneme sequence (Odden, 2013, p. 141).

Epenthesis

Epenthesis is the addition of sounds or elements of speech to a form that doesn't actually need it. In the *though* data that is pronounced as *though I have to*, students add phonological elements that are not found in the original lyrics.

According to Carr (2019, p. 115), epenthesis often appears as a strategy to facilitate speech production or maintain rhythm in speech.

Nasal Error

Nasal errors are found in the pronunciation *in my heart* to mean *my heart*. In this case, the nasal alveolar phoneme /n/ changes to the nasal bilabial phoneme /m/.

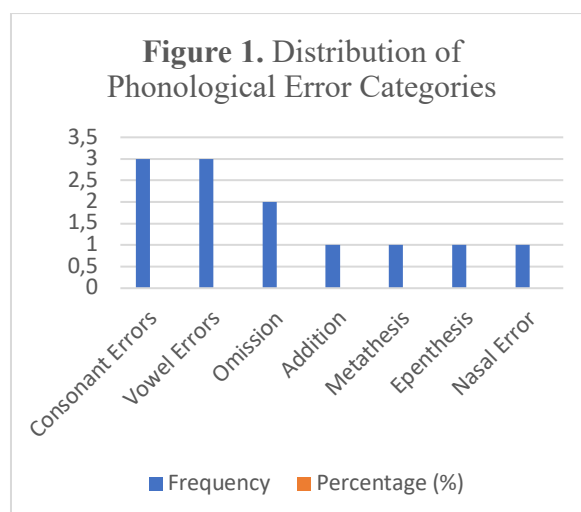
These changes indicate the assimilation of the articulation site which causes the nasal sound to move to a different articulatory position (Roach, 2009, p. 47).

Based on these findings, it can be concluded that phonological errors that appear in student singing include various segmental processes, both involving consonants and vocals. This variation of error suggests that students still face difficulties in producing sounds in English that are different from their first language phonological system.

Dominant Error Categories

Table 4. Frequency of Phonological Error Categories

Error Type	Frequency	Percentage (%)
Consonant Errors	3	25.0
Vowel Errors	3	25.0
Omission	2	16.67
Addition	1	8.33
Metathesis	1	8.33
Epenthesis	1	8.33
Nasal Error	1	8.33
Total	12	100



Based on Table 4 and Figure 1, the two most dominant categories of errors are consonant errors and vowel errors. Each category appeared three times or 25% of the total 12 cases analyzed.

The dominance of consonant errors can be seen in the case of replacing the /f/ sound with /p/, replacing /θ/ with /t/, and adding the phoneme /b/ in some words. These findings show that students have difficulty in producing English consonant sounds that are not found in the Indonesian phonology system. According to Avery and Ehrlich (1992, p. 15), second language learners tend to replace unfamiliar sounds with sounds that are closest to their first language.

Vowel errors also appear with the same frequency. These errors include vowel shift, vowel fusion, and monophthongization. The findings show that students still have difficulty in distinguishing the quality of English vowels, which are relatively more complex than the Indonesian vowel system. Roach (2009, p. 24) explained that English has much more vowel variations than many other languages so it often causes mistakes in foreign language learners.

Meanwhile, the omission category ranks next with a frequency of two cases or 16.67%. The addition, metathesis, epenthesis, and nasal error were only found once or 8.33% each.

Overall, the results showed that errors at the segmental level, particularly consonants and vowels, were the most dominant types of errors in student singing. These findings indicate that the production of English phonemes is still a major challenge for EFL students, especially when they have to imitate the pronunciation of English songs.

CONCLUSION

Based on the results of an analysis of 19 recordings of EFL students' singing using automatic transcription technology and IPA-based phonetic verification, this study found that not all detected irregularities can be categorized as pure phonological errors. After the selection and verification process, 12 cases were obtained that were considered representative to be further analyzed as phonological errors.

The results of the study showed that the phonological errors that appeared included seven main categories, namely **consonant errors, vowel errors, omissions, additions, metathesis, epenthesis, and nasal errors**. These errors reflect the difficulty students have in producing English sounds that differ from the Indonesian phonological system, especially in phonemes that do not have a direct equivalent in their first language.

Based on the frequency of occurrence, **consonant errors and vowel errors were the most dominant categories**, each with three cases (25%). Consonant errors are characterized by the substitution of sounds such as /f/ to /p/ and /θ/ to /t/, as well as the addition of certain consonants. Meanwhile, vowel errors appear in the form of vowel shift, vowel fusion, and

monophthongization. These findings suggest that segmental phoneme production is still a major challenge for EFL students when singing English songs.

In addition, this study found that students' pronunciation difficulties are influenced by several main factors, namely **L1 interference, limited phonological awareness, difficulties in processing connected speech, and the influence of auditory imitation when listening to songs**. Among these factors, first language interference appears to be the most dominant factor as most of the errors found are related to sound substitution that is not present in the Indonesian phonological system.

Methodologically, this study also shows that the integration of **Automatic Speech Recognition (ASR)** technology with phonetic analysis based **on the International Phonetic Alphabet (IPA)** can be an effective approach to identify and map students' phonological errors more objectively and systematically. Therefore, this approach has the potential to be used as a diagnostic tool in pronunciation learning to help lecturers and students understand areas of pronunciation difficulties more accurately and design learning strategies that are more targeted.

REFERENCES

- Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford University Press.
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education.
- Carr, P. (2019). *English phonetics and phonology: An introduction* (3rd ed.). Wiley Blackwell.
- Das, P., Acharjee, K., Das, P., & Prasad, V. (2015). Voice recognition system: Speech-to-text. *Journal of Applied and Fundamental Sciences*, 1(2), 191–195.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.
- Fromkin, V., Rodman, R., & Hyams, N. (2018). *An introduction to language* (11th ed.). Cengage Learning.
- Gerosa, M., Giuliani, D., Narayanan, S., & Potamianos, A. (2009). A review of ASR technologies for children's speech. ICMI-MLM1'09 Workshop on Child, Computer and Interaction, 1–8.
- Kelly, G. (2000). *How to teach pronunciation*. Longman.
- KOSHY JOSEPH. (2023, September 16). COMMON PRONUNCIATION ERRORS AND HOW TO CORRECT THEM.
- Misa, M. (2024). Language learning: Enhancing students' pronunciation through English songs. *Sintaksis: Publikasi Para Ahli Bahasa dan Sastra Inggris*, 2(6), 151–159. <https://doi.org/10.61132/sintaksis.v2i6.1289>
- Odden, D. (2013). *Introducing phonology* (2nd ed.). Cambridge University Press.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.
- Salim, R. F. Y. (2025). The effectiveness of English songs on students' pronunciation of eleventh grade students' of Swadhipa Natar High School South Lampung [Thesis, UIN Raden Intan Lampung].
- Shen, C. (2009). Using English songs: An enjoyable and effective approach to ELT. *English Language Teaching*, 2(1), 88–94.
- Yule, G. (2020). *The study of language* (7th ed.). Cambridge University Press.