

**PHONOLOGICAL ANALYSIS OF SEGMENTAL SPEECH SOUND
ERRORS OF INDONESIAN ONLINE ENGLISH TEACHERS**

THESIS

By:

Refi Ananda

NIM 220302110175



DEPARTMENT OF ENGLISH LITERATURE

FACULTY OF HUMANITIES

UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM

MALANG

2026

**PHONOLOGICAL ANALYSIS OF SEGMENTAL SPEECH SOUND
ERRORS OF INDONESIAN ONLINE ENGLISH TEACHERS**

THESIS

Presented to

Universitas Islam Negeri Maulana Malik Ibrahim Malang

In Partial Fulfillment of the Requirements for the Degree of *Sarjana Sastra* (S.S.)

By:

Refi Ananda

NIM 220302110175

Advisor:

Prof. Dr. Rohmani Nur Indah, M.Pd.

NIP 197609102003122011



DEPARTMENT OF ENGLISH LITERATURE

FACULTY OF HUMANITIES

UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM

MALANG

2026

STATEMENT OF AUTHORSHIP

I state that the thesis entitled “**Phonological Analysis of Segmental Speech Sound Errors of Indonesian Online English Teachers**” is my original work. I do not include any materials previously written or published by another person, except those cited as references and written in the bibliography. Hereby, if there is any objection or claim, I am the only person who is responsible for that.

Malang, 16 May 2026

The Researcher

Refi Ananda

NIM. 220302110175

APPROVAL SHEET

This is to certify Refi Ananda's thesis entitled **Phonological Analysis of Segmental Speech Sound Errors of Indonesian Online English Teachers** has been approved for thesis examination at the Faculty of Humanities, Universitas Islam Negeri Maulana Malik Ibrahim Malang, as one of the requirements for the degree of *Sarjana Sastra* (S.S.).

Malang, May 18, 2026

Approved by

Advisor,

Head of Department of English Literature,

Prof. Dr. Rohmani Nur Indah, M.Pd.

Dr. Agwin Degaf, M.A

NIP 197609102003122011

NIP 198805232015031004

Acknowledged by

Dean of Faculty of Humanities,

Dr. M. Faisol, M.Ag.

NIP. 197411012003121003

LEGITIMATION SHEET

This is to certify that Refi Ananda's thesis entitled **Phonological Analysis of Segmental Speech Sound Errors of Indonesian Online English Teachers** has been approved by the Board of Examiners as one of the requirements for the degree of *Sarjana Sastra* (S.S.) in the Department of English Literature.

Malang, May 18 2026

Board of Examiners

Signatures

Chair

1. Deny Efita Nur Rakhmawati, M.Pd.
NIP 198505302009122006

First Examiner

2. Prof. Dr. Rohmani Nur Indah, M.Pd.
NIP 197609102003122011

Second Examiner

3. Vita Nur Santi, M.Pd.
NIP 198306192011012008

Approved by

Dean of Faculty of Humanities

Dr. M. Faisol, M.Ag.

NIP 197411012003121003

MOTTO

إِنَّ مَعَ الْعُسْرِ يُسْرًا

“Indeed, after every difficulty comes ease.”

(Q.S. Al-Insyirah: 6)

DEDICATION

This thesis is dedicated to my beloved family Mr. Sarianto and Mrs. Artini, who have always been a source of strength and constant prayer for me and to Mrs. Sariati, who has given me unceasing love, prayers, and guidance. I also dedicate this work to my friends, who have always provided support and motivation throughout the process of writing this thesis.

ACKNOWLEDGEMENT

First and foremost, the researcher expresses her deepest gratitude and praise to Allah SWT, the Most Merciful and Compassionate, for His blessings, mercy, and grace, which have enabled him to complete his thesis titled “Phonological Analysis of Speech Sounds Error of Indonesian Online English Teachers Using Automatic Speech Recognition.” She also extends gratitude, blessings, and greetings to the Prophet Muhammad SAW, Allah’s messenger who serves as the best example for all of humanity.

The author realizes that this thesis could not have been completed without the help, support, and guidance provided by various individuals. Therefore, the author would like to express his deepest gratitude to Prof. Dr. Rohmani Nur Indah, M.Pd., for the time, guidance, and knowledge that he has shared with the author since the beginning of the writing process.

The researcher also expresses appreciation to the Dean and all faculty members of the Faculty of Humanities for their administrative support and for providing a conducive learning environment. Special thanks are also extended to all lecturers in the Department of English Literature who have shared their knowledge and insights from the first through the eighth semesters.

I would also like to express my deepest gratitude to my beloved family, my biological father Sarianto, my biological mother Artini, my beautiful mother Sariati and my brother Andika Khoirul Nizam. I want to convey my deepest gratitude for the love, care, and prayers you have given me. I am so grateful to be surrounded by such a loving family as you.

And to all my friends who have been with me from the start of my studies through the process of writing this thesis. To Reriskha Maharani, Thank you for fighting alongside me, for making sacrifices together, and for strengthening one another. Almas El Madina Aisyah and Ragil Putri Prasetyo, I would like to express my deepest gratitude for accompanying me throughout the writing process. And to Keysya Valentina Aulia, I would like to express my deepest gratitude for always being there for me in any situation and for always supporting me. I am deeply grateful to have found friendships that have brought me so many benefits.

The researcher acknowledges that this thesis is far from perfect. Therefore, feedback and suggestions from readers are greatly appreciated to help improve

this work in the future. The researcher hopes that this study will be beneficial to all, especially those interested in phonological errors.

Malang, 16 May 2026

The researcher,

Refi Ananda

NIM : 220302110175

ABSTRACT

Refi Ananda. (2026). *Phonological Speech Sound Errors Produced by Indonesian Online Non-Native English Teachers: An Analysis Using Automatic Speech Recognition (ASR)*. Undergraduate Thesis. Department of English Literature, Faculty of Humanities, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: Prof. Dr. Rohmani Nur Indah, M.Pd.

Key Terms: Phonological errors, Articulatory features, Pronunciation, Non-native English Teachers, Automatic Speech Recognition.

Many Indonesian non-native English teachers are doing English learning through online tools like YouTube; however, errors in pronunciations are quite common among them during the teaching process. Most of the past researches have concentrated on pronunciation errors made by the students or by the EFL learners while few studies have addressed phonological speech sound errors made by online non-native English teachers. Thus, this study intends to find out the types of pronunciation errors made by Indonesian online non-native English teachers and explains how the phonological speech sound errors are realized from an articulatory feature perspective. In this study, a qualitative descriptive approach was adopted, using phonetics and phonology approach with regard to Roach's theory (2009). The data used for this study were collected from the YouTube videos made by Indonesian online non-native English teachers using ASR transcription, manual verification, and IPA. Three major types of pronunciation errors have been found; they include substitution, addition, and omission with the substitution being the dominant error type. The errors committed show articulatory modifications that include place of articulation, manner of articulation, voicing, vowel quality, simplification of diphthong and consonant clusters. These findings indicate that pronunciation errors produced by Indonesian non-native English teachers are systematic and influenced by differences between the English phonological system and the speakers' first language. Future researchers are recommended to conduct more comprehensive studies involving different linguistic backgrounds and combining phonological analysis with interviews or observations.

ABSTRAK

Refi Ananda. (2026). *Phonological Speech Sound Errors Produced by Indonesian Online Non-Native English Teachers: An Analysis Using Automatic Speech Recognition (ASR)*. Undergraduate Thesis. Department of English Literature, Faculty of Humanities, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: Prof. Dr. Rohmani Nur Indah, M.Pd.

Key Terms: Phonological errors, Articulatory features, Pronunciation, Non-native English Teachers, Automatic Speech Recognition.

Banyak guru bahasa Inggris non-penutur asli di Indonesia yang mengajarkan bahasa Inggris melalui platform daring seperti YouTube; namun, kesalahan pengucapan cukup sering terjadi di antara mereka selama proses pengajaran. Sebagian besar penelitian sebelumnya berfokus pada kesalahan pengucapan yang dilakukan oleh siswa atau pelajar bahasa Inggris sebagai bahasa asing (EFL), sementara hanya sedikit penelitian yang membahas kesalahan bunyi fonologis yang dilakukan oleh guru bahasa Inggris non-penutur asli daring. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi jenis-jenis kesalahan pengucapan yang dilakukan oleh guru bahasa Inggris non-penutur asli daring di Indonesia dan menjelaskan bagaimana kesalahan bunyi fonologis tersebut terealisasi dari perspektif fitur artikulatoris. Dalam penelitian ini, pendekatan deskriptif kualitatif diadopsi, menggunakan pendekatan fonetik dan fonologi berdasarkan teori Roach (2009). Data yang digunakan dalam penelitian ini dikumpulkan dari video YouTube yang dibuat oleh guru bahasa Inggris non-penutur asli Indonesia secara daring menggunakan transkripsi ASR, verifikasi manual, dan IPA. Tiga jenis utama kesalahan pengucapan telah ditemukan; yaitu substitusi, penambahan, dan penghilangan, dengan substitusi sebagai jenis kesalahan yang dominan. Kesalahan yang dilakukan menunjukkan modifikasi artikulatoris yang meliputi tempat artikulasi, cara artikulasi, pengucapan bersuara, kualitas vokal, penyederhanaan diftong, dan gugus konsonan. Temuan ini menunjukkan bahwa kesalahan pengucapan yang dilakukan oleh guru bahasa Inggris non-penutur asli Indonesia bersifat sistematis dan dipengaruhi oleh perbedaan antara sistem fonologi bahasa Inggris dan bahasa pertama para penutur tersebut. Para peneliti di masa mendatang disarankan untuk melakukan studi yang lebih komprehensif dengan melibatkan latar belakang linguistik yang beragam serta menggabungkan analisis fonologi dengan wawancara atau pengamatan.

مستخلص البحث

ري في أناندا). 2026. (الأخطاء الصوتية في النطق التي يرتكبها مدرسو اللغة الإنجليزية غير الناطقين بها عبر الإنترنت في إندونيسيا: تحليل أطروحة. (ASR) باستخدام تقنية التعرف التلقائي على الكلام، البكالوريوس. قسم الأدب الإنجليزي، كلية العلوم الإنسانية، جامعة مولانا مالك إبراهيم الإسلامية الحكومية في مالانغ. المشرفة: الأستاذة الدكتورة روهما ني نور إنداه، ماجستير في التربية.

المصطلحات الرئيسية: الأخطاء الصوتية، السمات النطقية، النطق، مدرسو اللغة الإنجليزية غير الناطقين بها، التعرف التلقائي على الكلام.

يقوم العديد من مدرسي اللغة الإنجليزية غير الناطقين بها في إندونيسيا بتدريس اللغة الإنجليزية عبر منصات إلكترونية مثل يوتيوب؛ ومع ذلك، فإن أخطاء النطق تحدث بشكل متكرر بينهم أثناء عملية التدريس. وقد ركزت معظم الأبحاث السابقة على أخطاء النطق التي يرتكبها الطلاب في حين أن القليل فقط من الأبحاث تناولت الأخطاء، (EFL) أو متعلمو اللغة الإنجليزية كلغة أجنبية، الصوتية الفونولوجية التي يرتكبها مدرسو اللغة الإنجليزية غير الناطقين بها عبر الإنترنت. لذلك تهدف هذه الدراسة إلى تحديد أنواع أخطاء النطق التي يرتكبها معلمو اللغة الإنجليزية غير الناطقين بها عبر الإنترنت في إندونيسيا، وشرح كيفية حدوث هذه الأخطاء الصوتية الفونولوجية من منظور السمات النطقية. في هذه الدراسة، تم اعتماد نهج وصفي نوعي، باستخدام نهج صوتي وفونولوجي استنادًا إلى نظرية روتش (2009). تم جمع البيانات المستخدمة في هذه الدراسة من مقاطع فيديو على يوتيوب أنشأها مدرسو اللغة الإنجليزية غير الناطقين بها من إندونيسيا عبر الإنترنت، باستخدام تم تحديد ثلاثة أنواع رئيسية من أخطاء IPA والتحقق اليدوي ورموز (ASR) النسخ الصوتي الآلي النطق؛ وهي الاستبدال، والإضافة، والحذف، مع كون الاستبدال هو النوع السائد من الأخطاء. وتُظهر الأخطاء المرتكبة تعديلات في النطق تشمل مكان النطق، وطريقة النطق، والنطق الصوتي، ونوعية الأصوات، وتبسيط الأصوات المزدوجة، ومجموعات الحروف الساكنة. تشير هذه النتائج إلى أن أخطاء النطق التي يرتكبها مدرسو اللغة الإنجليزية غير الناطقين بها في إندونيسيا ذات طابع منهجي وتتأثر بالاختلافات بين النظام الصوتي للغة الإنجليزية واللغة الأم لهؤلاء المتحدثين. يُنصح الباحثون في المستقبل بإجراء دراسات أكثر شمولية تشمل خلفيات لغوية متنوعة، بالإضافة إلى دمج التحليل الصوتي مع المقابلات أو الملاحظة.

TABLE OF CONTENTS

STATEMENT OF AUTHORSHIP	iii
APPROVAL SHEET	iv
LEGITIMATION SHEET	v
MOTTO	vi
DEDICATION.....	vii
ACKNOWLEDGEMENT.....	viii
ABSTRACT.....	x
ABSTRAK	Error! Bookmark not defined.
مستخلص البحث	Error! Bookmark not defined.
CHAPTER I	1
INTRODUCTION.....	1
A. Background	1
B. Research Question	5
C. Significance of the Study	5
D. Scope and Limitation	5
E. Definition of Key Term	6
CHAPTER II.....	8
REVIEW OF RELATED LITERATURE.....	8
A. Phonology	8
B. Phonetics	9
C. Speech Sounds Error	11
D. Automatic Speech Recognition (ASR).....	12
CHAPTER III	15
RESEARCH METHOD	15

A. Research Design	15
B. Research Instrument	16
C. Data Source	16
D. Data & Data Collections.....	17
E. Data Analysis.....	18
CHAPTER IV.....	19
FINDING AND DISCUSSION	19
CHAPTER V	47
CONCLUSION AND SUGGESTION	47
REFERENCES.....	50
CURRICULUM VITAE.....	54
APPENDIX.....	0

CHAPTER I

INTRODUCTION

This chapter presents the background of the study which explain the importance of pronunciation in Indonesian online English language learning and the occurrence of speech sounds error produce by Indonesian online English teachers in YouTube-based learning videos. In addition, this chapter also explains the research questions, significance of the study, and scope and limitations of the research. At the end, there is also a definition of key terms.

A. Background

In this increasingly digital age, there have been significant changes in the world of education, including English language learning. Currently, many Indonesian English teachers conduct online lessons, such as using YouTube as a platform for delivering material. However, despite the use of media as a medium for delivering material, non-native English teachers are not immune to speech sounds error. According to Maxwell and Smith (2023), English learning videos delivered by teachers are not free from various mistakes.

One effective way to support pronunciation analysis is by using Automatic Speech Recognition (ASR) as a transcription tool. ASR helps convert spoken language into written text so that researchers can analyze pronunciation more systematically. In addition, teachers become more careful when delivering topics. Analyzing teachers' speech sounds error using ASR provides important insights into linguistic accuracy and pedagogical quality, especially since online English teachers often deal with diverse

linguistic backgrounds and technological limitations (Sun, 2023). Therefore, this issue highlights the importance of using ASR technology to systematically analyze speech sounds error made by Indonesian online English teachers.

Speech sounds error made by Indonesian online English teachers in this study will be analyzed using a phonetic approach. Phonetics is a branch of linguistics that studies how language or sounds are produced and articulated by speakers. According to Zalha and Lefiandi (2025), phonetic analysis is quite effective in identifying errors in the pronunciation of English consonants. Through a phonetic approach as a part of phonological study, this study focuses on analyzing segmental errors, especially in vowels or consonants, based on the results of automatic speech recognition (ASR) transcripts and careful listening. Therefore, the phonetic approach is suitable for analyzing speech sounds error in the context of online learning.

Phonology is a branch of linguistics that studies the sound system of a language and how those sounds are used and organized to distinguish meaning. It emphasizes sound patterns, sound distribution, and the rules that govern the use of sounds in a language system (Roach, 2009). Therefore, phonology not only discusses how sounds can be produced, but also how they function in the structure of language. Roach (2009) states that phonology is concerned with how sounds in a language can be organized into a system and how these sounds function to distinguish words and meanings. In English, the study of phonology covers segmental sounds, namely consonants and vowels, as well as the relationships and patterns of occurrence of these sounds in speech.

Pronunciation errors in education, especially among online English teachers, offers a significant topic to analyze. Furthermore, a comprehensive review of the development of research on pronunciation remarks the use of ASR in English as a foreign language

(EFL) (Liu, et al., 2025). In this study, the YouTube account of an English teacher who presents online English lessons is chosen to investigate the pronunciation errors. Based on preliminary observation, the learning content provided often contains errors when the teacher delivers English material. Not many researchers concern teachers' pronunciation as research subject, which makes this research significant and innovative compared to previous studies.

Previous studies have demonstrated the effectiveness of Automatic Speech Recognition (ASR) in analyzing and improving pronunciation among non-native English speakers. Liu et al. (2025) proved that ASR provides reliable, data-driven pronunciation evaluation, while Maxwell et al. (2023) showed that transfer learning enables ASR to detect pronunciation errors in bilingual contexts such as Indonesian-English. Ngo et al. (2024) further show that ASR significantly improves pronunciation accuracy in ESL and EFL learners, making it a practical feedback tool in language learning. Similarly, Spring and Tabuchi (2021) and Wei et al. (2022) highlight the practicality of ASR in online learning and its ability to effectively transcript the videos, despite some technical limitations. Supporting this claim, Dai and Wu (2023) and Dillon and Wells (2023) reveal that ASR-based pronunciation training provides instant corrective feedback, improving accuracy and confidence among learners. Overall, these studies highlight the growing pedagogical and analytical value of ASR in the field of pronunciation studies.

Based on the importance to analyze speech sounds error through ASR, Farrús (2023) emphasizes that ASR not only supports pronunciation improvement but also encourages independent learning, which can be applied to teachers as reflective practitioners in an online environment. In the phonetic approach, speech sound errors can occur when the sounds pronounced do not match the English pronunciation

standards. Sun (2023) further shows that pedagogy integrated with ASR improves speaking performance through technology-based learning. By combining these insights, this study fills a gap in existing research by applying ASR to analyze pronunciation errors among online English teachers, a group that is rarely examined despite their essential role as pronunciation models. Through a combination of phonetic approaches and the use of ASR as a tool for analysis, this study contributes to a deeper understanding of pronunciation errors made by online English teachers.

Further research is needed to understand how English teachers make pronunciation mistakes online via YouTube. This phenomenon is also very relevant in today's digital age, where the tools used for analysis are constantly evolving. By using ASR (Automatic Speech Recognition), it is possible to analyze and evaluate teachers' pronunciation mistakes. A word or sentence spoken by a teacher will serve as an example for their students, so if the teacher makes a mistake, students will misunderstand the pronunciation of the words knowledge being conveyed. For example, the pronunciation error of the word think /θɪŋk/ is often pronounced as tink /tɪŋk/.

This study is based on the assumption that English language learning delivered by teachers online can lead to pronunciation errors when delivering material. A further assumption is to analyze how pronunciation errors occurred in the speech of online English teachers using ASR tools. By analyzing these errors in greater depth using ASR, this research can also serve as a learning tool for online English teachers to correct and avoid pronunciation errors when teaching. This is done to reduce the negative impact on student comprehension. In addition, this is also beneficial for other online English teachers, encouraging them to be more careful in their English pronunciation when teaching online. If mistakes are not analyzed and evaluated, they will be repeated

and become difficult to correct. Therefore, it is important to analyze pronunciation mistakes using advanced technology in accordance with the correct methods.

B. Research Question

Based on the background of the study comprising the need to evaluate teacher pronunciation, the formulation of research questions is as follow:

1. What types of phonological speech sound errors are produced by Indonesian online English teachers?
2. How are the speech errors realized in terms of segmental features?

C. Significance of the Study

This study is expected to benefit readers both theoretically and practically in the field of phonetic analysis conducted on non-native English online teachers. Theoretically, this study contributes to phonetic studies, especially regarding sound errors that do not conform to English standards. The results of this study are expected to enrich the understanding of the types and characteristics of errors, both consonants and vowels, especially in the context of online learning through the YouTube platform.

In practical terms, this research can benefit non-native online English teachers as evaluation or learning material to be applied when teaching online. After identifying the types of errors that occur, teachers will be more careful in delivering material to students. Students will also be able to understand the material well without any misunderstandings in the process of receiving the material delivered by the teacher.

D. Scope and Limitation

This study focuses on analyzing pronunciation errors made by non-native English teachers who teach online through digital platforms, such as YouTube. This analysis

focuses speech sounds error made by Indonesian teachers during online teaching. To analyze the data, this study uses automatic speech recognition (ASR) as a tool to transcribe learning videos that have been uploaded to YouTube. This research is grounded in the theory of phonetics as a part of a phonology study proposed by Roach (2009), which focuses on the articulatory characteristics of English consonants and vowels and their representation using the International Phonetic Alphabet (IPA).

The limitations of this study are that the data used was only taken from a few videos of non-native English teachers on YouTube with a certain duration and number that were considered representative. This study does not address nonverbal aspects such as intonation, facial expressions, or body movements, but focuses only on pronunciation errors at the segment level (vowels and consonants). Furthermore, this study does not aim to assess the overall pedagogical abilities of teachers, but rather to analyze pronunciation errors from a linguistic and phonetic perspective using ASR technology as an analytical tool.

E. Definition of Key Term

- 1. Phonetics:** is a branch of linguistics that studies how sounds are produced, transmitted, and physically received by viewers which include segmental sounds such as vowels and consonants, including the differences in articulation of each sound produced.
- 2. Speech Sound Errors:** are the inaccuracies that occurred when Indonesian English teachers mispronounce a specific language sound (consonant or vowel), either by adding, changing, or omitting sounds.
- 3. Automatic Speech Recognition (ASR):** Automatic Speech Recognition (ASR) is a technology used to automatically convert spoken language into written text. In this

study, ASR is used only to assist the transcription process before the researcher conducts phonetic analysis.

4. Segmental features: are individual speech sounds, including vowels and consonants, that can be analyzed separately based on their articulatory characteristics. In this study, segmental features refer to the vowel and consonant sounds in English that are used to identify and explain pronunciation errors made by Indonesian online English teachers.

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter contains the theory needed to understand how speech sounds error produce by Indonesian online English teachers. Therefore, this chapter discusses phonology, phonetics, speech sounds error and automatic speech recognition (ASR).

A. Phonology

Phonology is a branch of linguistics that studies the system of sounds in a language and how those sounds are organized and used to distinguish meaning. Phonology does not discuss sounds as physical phenomena, but rather as part of a language system that has certain rules and sound patterns. Therefore, phonology focuses on the function of sounds in language, not on the process of articulation (Yule, 2010). Phonology discusses how the sounds of language can be organized in the minds of speakers and how these sounds can form recognizable patterns in a language. Each language has a different phonological system, so sounds that are considered correct in one language may not necessarily apply in another language. These differences in phonological systems are often a source of difficulty for foreign language speakers in learning the pronunciation of the target language (Fromkin et al., 2018).

Phonology discusses the rules that govern the occurrence and distribution of sounds in a language. These phonological rules can determine how sounds are pronounced in certain contexts and how these sounds interact with each other in a word or utterance. Therefore, pronunciation errors made by non-native speakers are often systematic because they are influenced by the phonological rules of their native language (Odden, 2005).

In the context of learning English as a foreign language, the study of phonology is very important for understanding the patterns of pronunciation errors that occur in non-native speakers. Differences between the phonological systems of English and the speakers' native languages can cause sounds to be substituted, omitted, or added. Therefore, phonology is used in this study as a theoretical basis for systematically explaining patterns of pronunciation errors, which can then be further analyzed through a phonetic approach (Yule, 2010).

Phonology focuses on the sound system of a language and the rules that govern how sounds are used and patterned. Differences between the phonological system of English and the speakers' first language often cause systematic pronunciation errors such as sound substitution, omission, or addition. Therefore, phonology is used in this study as a theoretical basis to explain patterns of pronunciation errors before they are further analyzed using a phonetic approach.

B. Phonetics

A branch of linguistics that discusses how sounds are obtained and produced, and how they are received by listeners. Phonetic studies do not discuss the meaning of a sound in a language system, but rather emphasize the articulation and physical aspects of spoken sounds. Thus, phonetics focuses on how the human vocal apparatus works and produces language sounds. Phonetic studies focus on how sounds are obtained, produced, and heard in their entirety by listeners in accordance with existing language standards. Thus, phonetics focuses on how the human vocal apparatus works to produce language sounds (Roach, 2009). Every language sound has certain articulatory characteristics that distinguish it from other language sounds. Therefore, phonetic

analysis has the potential to detect differences and identify language sounds and their places of articulation.

Phonetic analysis is used to help explain how sounds are pronounced and transcribed systematically using symbols from the International Phonetic Alphabet (IPA). Transcriptions in phonetic studies using IPA make it easier for researchers to record errors in detail and accurately. This is very important in research that analyzes errors in pronunciation, as it can be adjusted to existing English language standards (Roach, 2009). Phonetics is used to analyze segmental sounds such as vowels and consonants, including how and where each sound is articulated (Yule, 2020).

In the context of English, phonetic studies are used to analyze segmental sounds, namely vowels and consonants. Vowel sounds can be distinguished by the position of the tongue, lips, and the duration of a sound, while consonant sounds can be distinguished according to the method of articulation, place of articulation, and vibration of the vocal cords (Roach, 2009). Differences in sound systems between languages and between a speaker's first language and another language can cause errors in pronunciation, especially among non-native speakers. Therefore, this approach is suitable for analyzing errors in the pronunciation of English by non-native speakers.

Phonetics focuses on the physical and articulatory aspects of speech sounds and how these sounds are produced and perceived by listeners. By using phonetic analysis and IPA transcription, researchers can identify pronunciation errors accurately based on standard English pronunciation. Since English has different vowel and consonant systems from many other languages, non-native speakers often experience pronunciation difficulties. Therefore, a phonetic approach is appropriate for analyzing pronunciation errors made by non-native English speakers in this study.

C. Speech Sounds Error

Speech sound errors are a form of discrepancy between actual speech and existing English language standards. In phonetic studies, pronunciation errors are analyzed in terms of physical sound realization, particularly in segmental sounds, namely vowels and consonants. This analysis is carried out by comparing the pronunciation produced by the speaker or sound producer, which is then represented using International Phonetic Alphabet (IPA) symbols. Speech sounds error can occur when the speech produced by the speaker does not match the actual articulatory form (Roach, 2009). In Indonesian online English teachers, sound errors can occur due to differences in articulation between standard English sounds and the non-native speaker's first language. As a result, speakers often replace English articulation with sounds from their first language that they consider to be similar or the same.

The most common types of speech sounds error in phonetic studies are sound substitution, sound addition, and sound deletion. Sound deletion occurs when a speaker does not pronounce a sound as it should be according to the existing language standard. For instance, a speaker may omit a final consonant sound in an English word, which changes the pronunciation of the word and may affect intelligibility. Sound substitution occurs when a sound is replaced by another sound. For example, in pronunciation of Indonesian English teachers may replace the dental fricative sound /θ/ with an alveolar sound /t/ or /s/ because the dental fricative sound is not present in their native sound system. Meanwhile, sound addition occurs when a speaker adds an extra sound that should not be present in the standard pronunciation of a word. This type of error usually happens when speakers insert an additional vowel or consonant to make pronunciation easier, especially in words with consonant clusters. For example, a speaker may add a vowel sound between two consonants to avoid difficult pronunciation (Roach, 2009).

In the context of English language learning, pronunciation errors can play an important role as a learning model for students. Mistakes made by teachers can have fatal consequences on students' understanding of the material that should be learned. Pronunciation should be in accordance with existing language standards so that there is no misunderstanding of the meaning of the sound. Therefore, speech sounds error analysis can play an important role in identifying and evaluating mistakes made by teachers, especially in online learning for non-native English teachers.

Speech sounds errors occur when the sounds produced by speakers do not match standard English pronunciation. These errors commonly appear in segmental sounds, including vowels and consonants, and are influenced by differences between English sound articulation and the speaker's first language. Common types of speech sound errors include sound substitution, sound deletion, and sound addition. In the context of English language learning, especially in online instruction, teachers' speech sounds play an important role as a model for students. Therefore, analyzing speech sound errors is necessary to identify pronunciation mistakes and improve the quality of English teaching by non-native speakers.

D. Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) is a technology that converts speech into written text automatically. In linguistic research, ASR is commonly used as a transcription tool to assist researchers in documenting spoken data more efficiently. Research such as Spring & Tabuchi (2021) and Wei et al. (2022) shows that ASR is accurate in identifying pronunciation errors made by non-native speakers. Several studies show how Automatic Speech Recognition (ASR) can support pronunciation analysis among non-native English speakers. Recent research also proves that ASR is

effective for analyzing pronunciation, including Indonesian-English accents. Therefore, ASR is used in this study to objectively detect speech errors made by online English teachers. In linguistic research, ASR can be used as a tool to transcribe videos into text in order to identify errors made by a speaker. ASR can make it easier for researchers to transcribe videos and do so more quickly, even though manual checking is still required as a form of correction if there are errors and to ensure the accuracy of the transcription tools.

Automatic speech recognition is used to transcribe videos from YouTube about non-native English teachers. ASR offers several benefits in language research. It helps researchers transcribe speech data faster than manual transcription and allows large amounts of spoken data, such as online learning videos, to be documented efficiently. In pronunciation research, ASR transcripts can serve as initial references for researchers before conducting detailed phonetic analysis using listening and phonetic transcription such as the International Phonetic Alphabet (IPA). The data is then analyzed using phonetic studies, particularly on segmental sounds, vowels, and consonants, by comparing the actual sounds with the sounds according to existing language standards, using IPA symbols. Using phonetic studies and ASR as tools to transcribe video data from YouTube can make research more systematic and structured. For example, when a teacher pronounces the word *think* /θɪŋk/ as *tink* /tɪŋk/, ASR can capture this pronunciation in text form, which can then be analyzed phonemically to identify a sound substitution error. This process allows researchers to systematically detect speech sound errors in online teaching videos (Maxwell-Smith & Foley, 2023). The use of ASR technology in this study demonstrates increasingly sophisticated technological developments, particularly in the context of online learning. Therefore, ASR

technology is not merely a tool to assist in the data analysis process, but an important part of every research process.

Automatic Speech Recognition (ASR) is an effective tool for transcribing YouTube videos more quickly and systematically, while phonetic analysis using IPA allows the identification of errors in vowels and consonants based on standard English pronunciation. Although manual checking is still needed to ensure accuracy, the combination of ASR and phonetic analysis makes the research process more efficient and structured, especially in analyzing speech produced by Indonesian online English teachers. Therefore, researchers need to anticipate these limitations when using ASR. The transcription results should be manually reviewed and verified by listening to the original audio recordings. In this study, ASR is used only to assist the transcription process, while the identification and analysis of pronunciation errors are conducted manually by the researcher through phonetic analysis.

CHAPTER III

RESEARCH METHOD

This chapter explains the research methodology framework used in this study, first is explanation of the research design, data, and data sources. The next section also explains the procedures for data collection and data analysis techniques.

A. Research Design

This study used a qualitative descriptive research design, aiming to describe, explain, and identify the types of sound errors experienced by non-native English teachers in online learning. According to Creswell & Creswell (2018), qualitative research is an investigative approach used to understand social or humanitarian issues based on the meanings experienced by participants in their real contexts, where data is usually collected in the form of words rather than numbers, and researchers play a key role in data collection and analysis. Furthermore, Sugiyono (2019) explains that qualitative descriptive research is suitable for studies that aim to portray and interpret existing phenomena without manipulating variables.

Qualitative research in applied linguistics allows researchers to closely examine naturally occurring language data, such as spoken discourse, in order to identify patterns and features of language use (Mackey & Gass, 2016). The qualitative approach has the potential to provide more detailed explanations and in-depth analysis of the data obtained and the emergence of naturally spoken errors in the video. The qualitative approach used in the study to explain the results of the data analysis.

B. Research Instrument

In this qualitative research, the researcher played a key role in the research, responsible for data selection, speech error analysis, and interpretation of results. She determined relevant video segments, then inserted the videos into ASR and identified errors using a phonetics approach. Several supporting tools were used as follows: ASR was used to generate automatic transcriptions of spoken data from YouTube videos. The transcription results were then verified manually by the researcher before further phonetic analysis for categorizing errors based on their type (e.g., substitution, addition, omission, and segment errors). YouTube downloader was used to download selected YouTube videos, making it easier to transcribe and identify errors in the data. These tools are used for data collection and help clarify the data analysis process.

C. Data Source

This study consists of linguistic units containing speech sounds error. These errors include words, phrases, or sentences spoken by non-native online English teachers that deviate from standard English pronunciation. The data was obtained and then transcribed using ASR.

Data sources for this study came from videos uploaded to a digital platform (YouTube) containing English lessons taught by non-native English teachers. These videos were selected because they represent real teaching and learning situations where accuracy in pronunciation is very important for learners, there are two accounts of YouTube. Legibility criteria, the video must be presented by a non-native speaker teacher, the video must contain content related to English language instruction. The teacher must use spoken English as the medium for delivering the material. The audio quality must be clear enough to be easily transcribed into ASR. The video content

consists of oral instruction, not numerous music clips. Meanwhile the exclusion criteria cover noisy videos that have distracting background sounds and when the teacher mixes languages excessively, making it difficult to detect errors. Another exclusion occurs when the video is too short, resulting in incomplete material delivery. The summary of the selection criteria is presented in table 1.

Table 1. Inclusion and exclusion criteria for video selection

Criteria Type	Inclusion Criteria	Exclusion Criteria
Teacher Identity	The video is presented by Indonesian non-native English teacher.	The video is presented by native English teacher.
Content Type	The video contains English language teaching or explanation English materials.	Video containing music or unrelated content of teaching.
Language Use	The teacher using English for explain the materials.	The teacher using Indonesian for explain the materials.
Video Duration	The duration of video is between 5-15 minutes.	Very shorter than 4 minutes.

Based on the selection criteria, the videos chosen are those as listed in table 2.

Table 2. Link data video from YouTube

Account profile	Video link
1.Juanita Melagros	http://www.youtube.com/@juanitamelagros
2.Naila Farhana	http://www.youtube.com/@nailafarhana

D. Data & Data Collections

Data collection was conducted in a clear and systematic way based on the research procedures. First, the researcher selects online English teaching videos by non-native speakers that meet the inclusion criteria. Next, the selected YouTube videos are downloaded using a video downloader and stored for the next process. Then, the videos are transcribed using automatic speech recognition (ASR) technology to identify the spoken text. After the transcription process, the researcher listens carefully to the videos

and compares the actual pronunciation with the standard English pronunciation. This step is conducted to identify pronunciation errors through phonetic analysis.

E. Data Analysis

The data in this study are analyzed qualitatively through several systematic stages. First, the selected YouTube videos are transcribed using Automatic Speech Recognition (ASR) to obtain an initial written representation of the spoken data. However, the analysis does not rely solely on the ASR output. The researcher manually rechecks the transcription by listening carefully to the original audio recordings to ensure the accuracy of the spoken data. Second, the researcher compares the teachers' pronunciation with standard English pronunciation based on phonetic transcription using the International Phonetic Alphabet (IPA). This step is conducted to identify pronunciation deviations that occur in vowels or consonants. Third, the identified pronunciation errors are categorized based on their types, such as sound substitution, sound addition, and sound omission. A classification sheet is used to maintain consistency in the categorization process.

Finally, the researcher interprets the patterns of pronunciation errors by referring to phonetic and phonological theories. The results of the analysis are then summarized to explain the types and characteristics of speech sound errors produced by Indonesian online English teachers. In this study, ASR functions only as a tool to assist the transcription process, while the identification and analysis of pronunciation errors are conducted manually.

CHAPTER IV

FINDING AND DISCUSSION

This chapter presents the research findings and relates them to the research questions. It focuses on the types of pronunciation errors made by Indonesian online English teachers in YouTube videos. The data were transcribed using Automatic Speech Recognition (ASR) technology and further analyzed through manual phonetic analysis based on the framework proposed by Peter Roach (2009) to provide a deeper understanding of the pronunciation errors identified in the data.

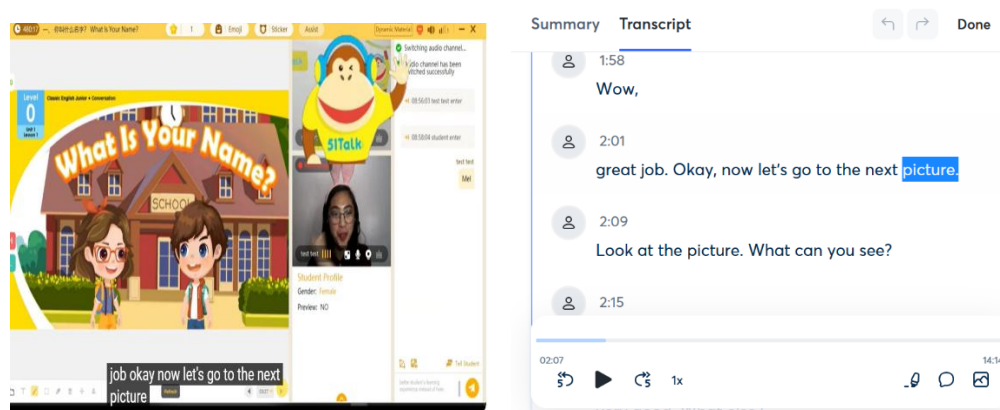
A. FINDINGS

In this chapter, the research findings are presented in the form of pronunciation errors made by Indonesian online English teachers in instructional videos on YouTube. Data were collected from two YouTube accounts belonging to teachers who are non-native English speakers and extracted from several selected videos; the data were then transcribed using Automatic Speech Recognition (ASR) technology and manually verified by the researchers to ensure their accuracy. Based on the collected data, several pronunciation errors were identified and classified into three types: substitution, addition, and omission. This classification is based on the phonetic framework proposed by Peter Roach (2009), which focuses on the articulatory characteristics of consonants and vowels in English.

After undergoing data reduction, the selected data is presented in the form of data points. Each data point represents one instance of a pronunciation error found in the teachers' speech, which is then analyzed based on the differences between standard pronunciation and actual pronunciation. The analysis is conducted in accordance with

the theory used, namely Peter Roach's theory (2009); the errors analyzed will be categorized based on the types of errors present. The pronunciation errors were taken from the YouTube channels of two non-native English teachers; therefore, each data point is labeled with a code. For example, (1.1) indicates that the data was obtained from Teacher 1's YouTube channel, and (2.1) indicates that the data was obtained from Teacher 2's YouTube channel.

Datum 1.1



The interaction depicted in Datum 1.1 is from a video on learning the English language on YouTube, in which the teacher gives a lesson to the students. In this case, the teacher instructs the students to proceed to the next stage of the lesson. The occurrence takes place in a learning context, which may be compared to a classroom scenario in which the teacher imparts the lesson in the English language. Throughout the entire duration of the lesson, the English language is predominant in the imparting of the lesson. However, although the teacher seems very sure of himself while giving the lesson, he pronounces some words incorrectly.

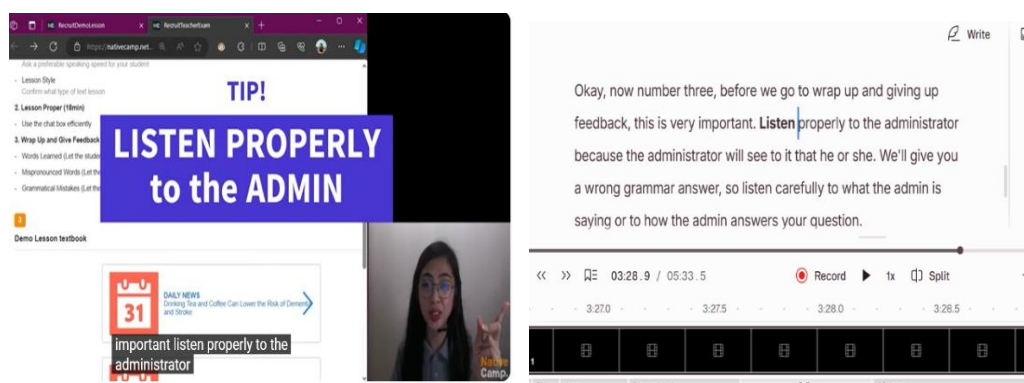
At this point, the teacher said, "Let us examine the following picture." In this phrase, one can identify a mistake in the pronunciation of the word "picture" as shown in the following audio clip:

https://drive.google.com/drive/folders/1CZaE2E3o4_BGUgqN5-AA99jfpHsvs2X-. In accordance with standard pronunciation rules, the word “picture” is pronounced as /'pɪktʃər/. Nevertheless, in this case, it sounded like /'fɪktʃər/

The phenomenon takes place as the /p/ consonant sound, found in an initial position in a word, is substituted for the /f/ sound. As stated by Roach (2009), the /p/ sound can be characterized as a voiceless bilabial stop. It is formed when both lips are pressed together, and then the air is expelled. In contrast, the /f/ sound is a voiceless labiodental fricative, which means that this sound is created when the lower lip is positioned against the upper teeth, thus allowing air to pass through the narrow space.

These errors can be seen through the difference in the places of articulation (bilabial vs. labiodental) and manners of articulation (plosive vs. fricative). Furthermore, errors made at the onset of a word are especially crucial since the onset usually provides a clue to the identity of a word. Besides, these errors could also be made due to the influence of the first language of the speaker, as some sounds do not exist in some languages. Thus, it would appear that this error is a substitution, where one sound substitutes another. It shows that there might still be mistakes in pronunciation even in a learning situation.

Datum 1.2



The screenshot displays a video lesson interface. On the left, a sidebar lists lesson steps: 1. Lesson Style, 2. Lesson Proper (10min), 3. Wrap Up and Give Feedback. A large blue box with white text reads "TIP! LISTEN PROPERLY to the ADMIN". Below this, a smaller box says "important listen properly to the administrator". The main video area shows a woman speaking. On the right, a transcript window displays the text: "Okay, now number three, before we go to wrap up and giving up feedback, this is very important. Listen properly to the administrator because the administrator will see to it that he or she. We'll give you a wrong grammar answer, so listen carefully to what the admin is saying or to how the admin answers your question." The transcript window also includes a timeline and playback controls.

Datum 1.2 depicts an interaction taken from an English language learning video on YouTube, in which a teacher gives instructions to students during a learning activity. In this context, the teacher asks the students to listen carefully to the explanation provided. The interaction takes place in an online learning environment, where English is used as the primary language for delivering the material. The teacher appears to speak quite fluently, although there are still some pronunciation errors on certain words.

In that interaction, the teacher said, “It’s important to listen carefully to the administrator.”(https://drive.google.com/drive/folders/1CZaE2E3o4_BGUgqN5-AA99jfpHsvs2X-) In this utterance, there is a pronunciation error in the word “listen.” In standard English pronunciation, the word “listen” is pronounced /'lɪsən/. However, the teacher pronounced it as /'lɪstən/, which indicates a deviation from standard pronunciation.

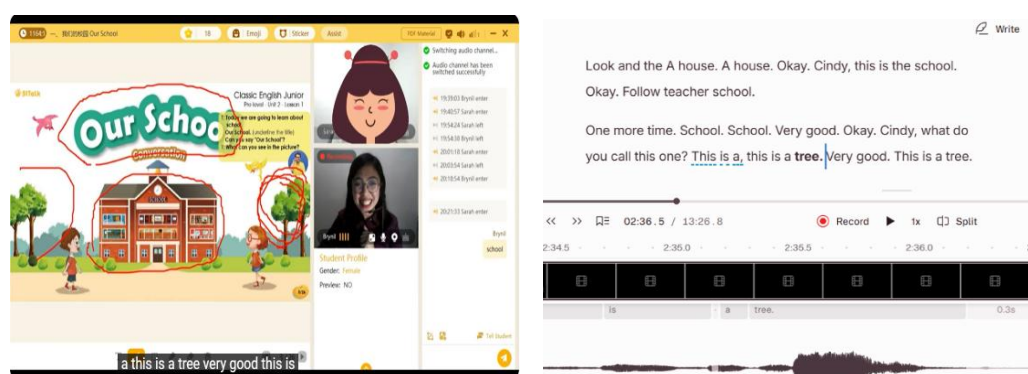
This deviation occurs due to the addition of the consonant sound /t/ in the middle of a word (medial consonant). In its standard form, the letter t in the word “listen” is a silent consonant that is not pronounced. According to Roach (2009), there are several words in English whose spelling does not always correspond to their pronunciation, so certain consonants may not be pronounced.

The /t/ sound itself is classified as a voiceless alveolar plosive, which is a sound produced by placing the tip of the tongue against the alveolar ridge (upper gum) and then suddenly releasing the air. In the word “listen,” this sound should not be pronounced. However, speakers still pronounce it, resulting in the form /'lɪstən/. This error indicates that the speaker is likely influenced by the written form of the word (orthographic influence), causing them to pronounce each letter in the word exactly as

it is written. This phenomenon is quite common among non-native speakers who still rely on spelling when pronouncing English vocabulary.

Therefore, this error is classified as an addition. It involves the insertion of a consonant sound in the middle of a word where none should be pronounced. This finding indicates that the discrepancy between spelling and pronunciation in English remains a challenge for non-native-speaking teachers when delivering oral instruction during the learning process.

Datum 1.3



Example 1.3 shows an interaction from a video about English language acquisition available on YouTube. In this case, the teacher explains vocabulary or refers to an object that should be explained to the students. In this particular case, the teacher provides the word “tree” as an example of the object that the class is working on. The interaction took place during an online class and, thus, the teacher used English as the primary language of explanation. It should be noted that the teacher delivered the explanation confidently; however, there were some inaccuracies concerning pronunciation of the words.

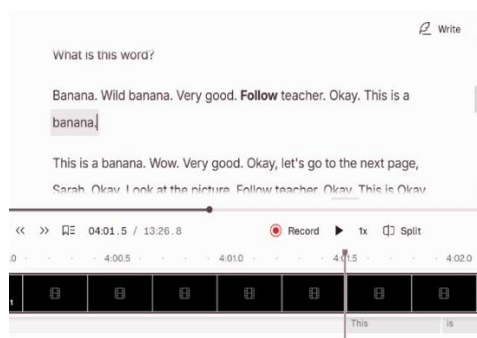
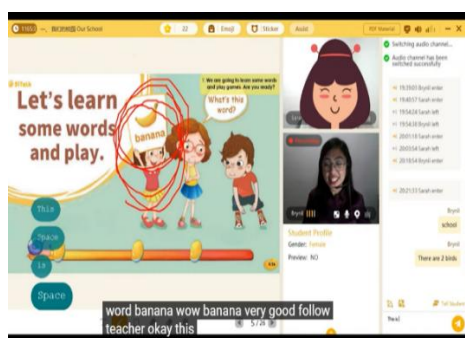
In that interaction, the teacher said, “This is a tree; it’s very nice.” Here, the pronunciation of the word “tree” was not accurate because according to the standard

pronunciation /tri:/. However, in this case, it sounded like /dʒri:/. It occurs because the consonant sound /t/ as the initial consonant, is replaced by the sound /dʒ/.

According to Roach (2009), the sound /t/ is a voiceless alveolar plosive, that is, a sound produced by placing the tip of the tongue against the alveolar ridge (upper gum) and then suddenly releasing air without vocal cord vibration. Meanwhile, the sound /dʒ/ is a voiced postalveolar affricate, a sound produced by blocking the airflow in the postalveolar region followed by friction and accompanied by vocal cord vibration. These changes reflect differences in place of articulation (from alveolar to postalveolar), manner of articulation (from plosive to fricative), and vowel status (from voiceless to voiced). Errors in the initial sound of a word are particularly significant because the initial sound serves as the primary cue for recognizing spoken words.

This mistake can be associated with the pronunciation skills of the speaker's native language or the inability to pronounce the cluster /tr/. As a consequence, the learner may replace /t/ by an alternative sound that suits him/her better, which gives the pronunciation variant /dʒri:/. This error is regarded as substitution since it includes replacement of one sound by another at the initial position of the consonant. The pronunciation of consonants' clusters at the beginning of words still poses some difficulties for English-speaking teachers.

Datum 1.4



In Datum 1.4, there is an example of an exchange that was taken from a video on YouTube about learning the English language. In the exchange, a teacher provides instructions to the learners in the learning process. In this exchange, a teacher tells a learner named Sarah to follow the instructions of the teacher while learning. The exchange happens in a virtual classroom whereby the teacher uses English as the medium of instruction to communicate with the learners. The teacher is highly competent in giving instructions despite the pronunciation errors.

During the encounter, the teacher uttered the sentence, “Sarah, follow the teacher.” In the sentence above, there is an incorrect pronunciation of the word “follow.” While the correct pronunciation of the word in British English is /'fɒləʊ/ and /'fɑ:lou/ in American English, the teacher pronounces the word as /'falou/, meaning “fallow.”

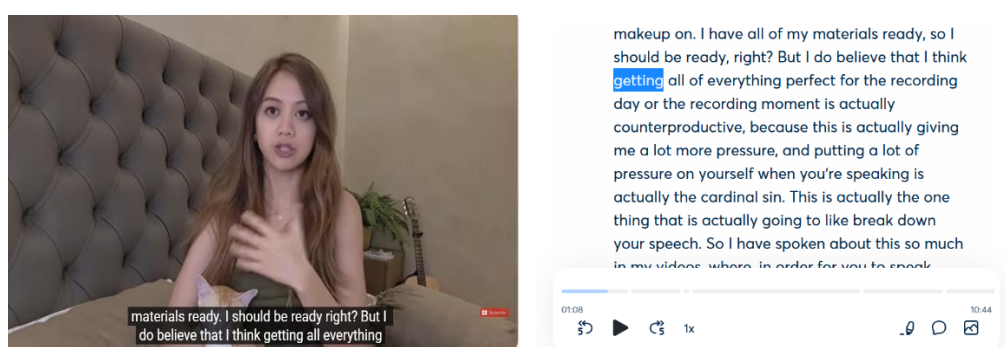
This alternation is caused by the substitution of the vowel phoneme /ɒ/ (or /ɑ:/ in the case of American English) that occupies the first syllable by the phoneme /æ/ or /a/. As noted by Roach (2009), the phoneme /ɒ/ is characterized by a round open back vowel, while the phoneme /æ/ represents an unround near-open front vowel. It is possible to see that there is a shift in the position of the tongue from the back to the front of the mouth.

This is an error that influences the quality of pronunciation of the vowel in the first syllable, leading to a difference in the pronunciation of the word ‘follow’. This is actually a serious error since changing the pronunciation of vowels might influence the meaning of the word entirely.

The reasons for such mistakes are probably caused by the fact that the speaker has a certain linguistic background from their first language. This is because it can be

very difficult for speakers to distinguish vowel sounds which do not exist in their native tongue, and therefore they replace them with familiar vowel sounds. This is why the mistake falls into the category of “substitution” since it implies the interchange of vowels occurring in the middle of a word, namely in its nucleus, the first syllable. In other words, discrepancies in the vowel systems of English and L1 are still an issue in the pronunciation of English words.

Datum 2.1



The example presented in Datum 2.1 refers to the dialogue between the teacher and students extracted from the English language learning video on YouTube. In this case, the speech of the teacher is spontaneous since it is an explanation of a particular idea, delivered in a natural way during the class. Moreover, English language serves as the medium of communication in this example. It should be emphasized that despite the fluency of the explanation, the speech contains mispronunciation of certain words.

The dialogue is characterized by mispronunciation of the word “getting.” In accordance with the rules of standard English pronunciation, this word should be pronounced in the following way: /'getɪŋ/. However, the teacher pronounced it as /'dʒetɪŋ/, which indicates a deviation from standard pronunciation. This deviation may occur because the consonant sound /g/, which occurs at the beginning of a word (initial consonant), is replaced by the sound /dʒ/. According to Roach (2009), the sound /g/ is

a voiced velar plosive, that is, a sound produced by raising the back of the tongue to the soft palate (velum) so that the airflow is briefly blocked, then released accompanied by vocal cord vibration. Meanwhile, the sound /dʒ/ is a voiced postalveolar affricate, which is a sound produced by blocking the airflow in the postalveolar region followed by friction and accompanied by vocal cord vibration.

This change indicates a shift in both the place of articulation (from velar to postalveolar) and the manner of articulation (from a plosive to an affricate). Errors in the initial sound of a word are particularly significant because the initial sound is one of the primary elements in recognizing spoken words. The change of the sound /g/ to /dʒ/ can cause the word to sound different from its standard form.

This mistake could either be affected by the speaker's first language background or by the inability to properly produce the sound /g/. In the process, the speaker replaces the sound with another sound which is perceived to be easier to pronounce. It belongs to a substitution because it entails substituting a certain sound with another sound at the beginning of a word. From these results, it can be seen that the pronunciation of velars might still remain difficult for certain non-native speakers who are teachers.

Datum 2.2

Datum 2.2 illustrates an exchange from a language learning video in English available on YouTube. The exchange involves a teacher teaching the content to the learners in a Q&A session. In this case, the teacher gives examples of how to respond to common, simple questions in daily conversations. The exchange occurs in an online learning platform, where English is the main language used by the teacher to teach the content. It should be noted that the teacher seems quite articulate while teaching the content but makes some pronunciation mistakes.

During the discussion, the teacher made use of the expression “to answer a simple question.” The pronunciation of the word “answer” is wrong in the phrase. The right pronunciation of the word in the English language is /'ænsər/, while the teacher used /'ænswər/.

This deviation occurs due to the addition of the consonant sound /w/ in the middle of a word (medial consonant). In standard spelling, the letter w in the word “answer” is a silent consonant that is not pronounced. According to Roach (2009), in English there are several words whose spelling does not always correspond to their pronunciation, so certain letters are not pronounced when the word is spoken.

Sound /w/ belongs to the category of voiced labio-velar approximants, which means that the articulation involves rounding the lips and placing the back part of the tongue close to the soft palate while the vocal cords vibrate. The occurrence of this sound in the word “answer” is not correct because it should not be present. However, the sound is pronounced, yielding /'ænswər/.

It is an example of the orthographic influence, as the pronunciation of the word “answer” corresponds to its spelling. It can be concluded that the person who produces such an articulation pronounces each letter in accordance with its spelling. It is evident that this is an error of addition because it consists of inserting a consonantal sound in the middle of a word where there should be no sound at all. It implies that the difference between the orthography and phonology of English continues to pose difficulties for teachers who speak English as their second language in teaching English pronunciation.

Datum 2.3

Datum 2.3 illustrates an interaction that comes from a language learning video on YouTube where a teacher is talking about something or giving her views to her

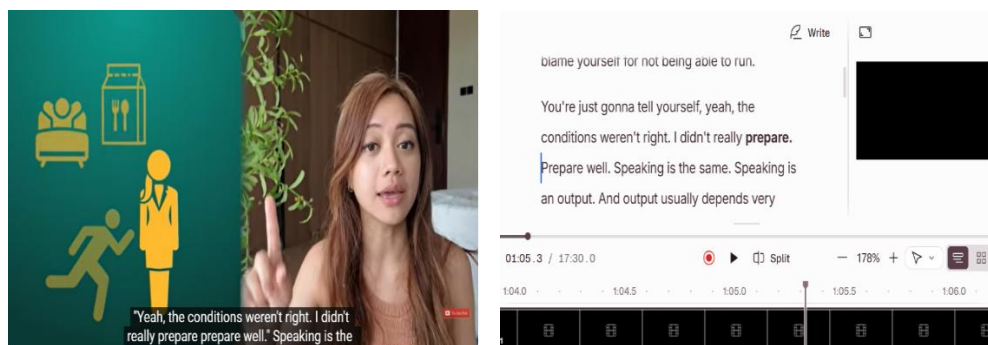
learners as part of the language learning process. In this case, the teacher talks spontaneously and communicates in English. It happens in an online learning environment where the teacher seems very fluent in giving her view although there are some errors in pronouncing some words.

In the mentioned dialogue, the teacher uttered, “right now, while I’m here.” There is mispronunciation of the word “while.” In general pronunciation of English words, the word “while” is pronounced /waɪl/. In the above case, it has been pronounced as /wel/.

This deviation occurs because the diphthong /aɪ/, which is a medial vowel, is replaced by the monophthong /e/. According to Roach (2009), the diphthong /aɪ/ is a combination of two vowel qualities that shift from an open vowel to a high front vowel. Meanwhile, /e/ is a front mid vowel that can be produced with the tongue positioned at the front of the oral cavity without any shift in sound quality.

This shows that the speaker has reduced the diphthong to a monophthong thus causing a change in the quality of the vowel. The error impacts on the pronouncing of the word as the word "while" appears different from the expected pronunciation and may cause confusion to other people. It can be caused by the first language interference as in many cases, there are no English diphthongs in some languages or where there are English diphthongs, they are not very clear. As a result, people end up substituting the English diphthongs for vowels that appear easy. This is a substitution error since the mistake entails changing one vowel for another in the middle of the word. From the above results, we conclude that producing diphthongs continues to pose challenges to foreign teachers in using English in oral communication.

Datum 2.4



Datum 2.4 depicts Datum 2.4 presents the excerpt from the English language learning video from YouTube, where the instructor provides their experience or situation to the students. In the particular case, the instructor conveys that they did not prepare adequately before carrying out the activity. The interaction happens in an online learning environment, where the medium of instruction is English. The speaker demonstrates adequate proficiency in speaking this utterance; however, there are still some mistakes in pronouncing specific words.

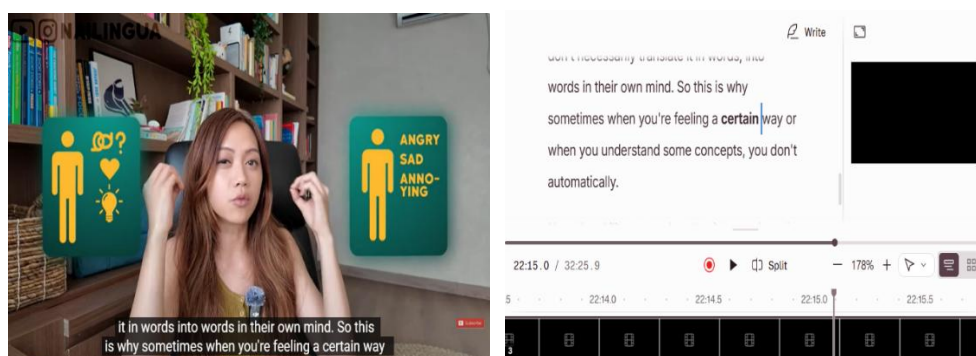
In the above statement, the instructor says, “I didn’t prepare well”. This statement involves a mistake in the pronunciation of the word “prepare.” According to the standard pronunciation of the word “prepare,” the word is pronounced as /prɪ'peə/ (British English) or /prɪ'per/ (American English). Nonetheless, the speaker pronounces the word as /pripar/.

This deviation may occur because the vowel sound /ɪ/ in the first syllable and the diphthong /eə/ (or /er/) in the second syllable have changed into the monophthongs /i/ and /a/. According to Roach (2009), the sound /ɪ/ is a near-close front unrounded vowel, while the diphthong /eə/ is a combination of two vowel qualities that may involve a shift in tongue position from a mid to a central vowel. In the speech produced

by speakers, this diphthong is simplified into the single vowel /a/, resulting in a significant change in vowel quality.

This change can be seen as an indication of tongue movement and properties of the vowel from front to an open and centralized vowel. This mispronunciation can impact the quality of the pronunciation as the word "prepare" is not uttered in its conventional manner and can cause misunderstanding among the hearers. It can be attributed to the influence of the first language of the speaker. In the first language, the system of English vowels, particularly diphthongs, does not always exist or have distinctions. Therefore, this error can be classified as a substitution, since one vowel sound is replaced by another within the same word. This finding indicates that the production of complex vowel sounds, such as diphthongs, remains a challenge for some non-native teachers in spoken English.

Datum 2.5



Datum 2.5 is an excerpt from the dialogue of a conversation on YouTube video designed for the purpose of teaching English, where a teacher describes an emotion or feeling. In this scenario, the teacher uses the word “certain” in his/her explanation about the emotion or feeling that one might experience in the particular situation. The conversation is taking place in an online classroom environment in which English is spoken.

In the conversation above, the teacher utters the sentence, “when you’re feeling a certain way.” Here, there is an error in the pronunciation of the word “certain.” In English pronunciation, the word “certain” is said as /'sɜ:tən/. Here, the teacher deviates from the standard pronunciation as he/she pronounces the word as /'sɜ:em/.

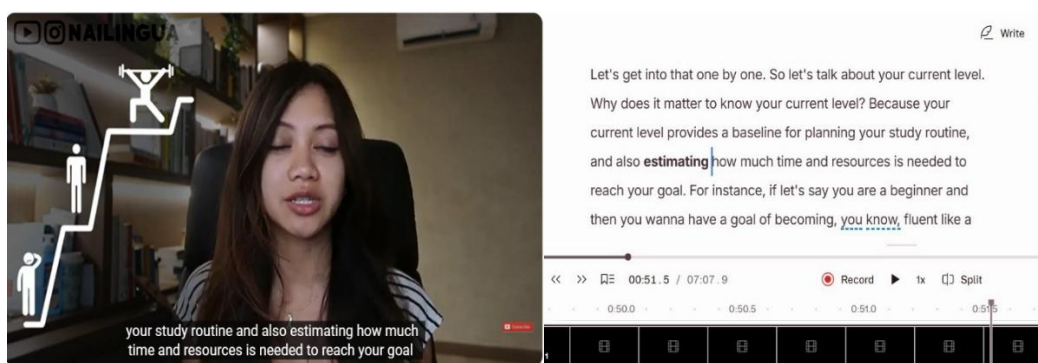
This deviation can occur due to two types of phonological errors. First, there is the omission of the consonant /t/ in the middle of a word (medial consonant). According to Peter Roach (2009), sound omission occurs when a sound that should be pronounced is not actually produced, thereby altering the phonological structure of the word. The /t/ sound in this word is a voiceless alveolar plosive, which is a sound produced by placing the tip of the tongue on the alveolar ridge and then releasing the air suddenly. However, the speaker does not pronounce this sound.

Next, there is a substitution of the vowel sound /ə/ for the diphthong /eɪ/. According to Roach (2009), the sound /ə/ (schwa) is a mid vowel that typically occurs in an unstressed syllable, whereas /eɪ/ is a diphthong that can involve a shift from a mid vowel to a high front vowel. This change may indicate a shift in vowel quality from the expected form to another form that is different in terms of articulation.

The changes can be made to either the consonants or the vowels in the word which results in an utterance that is considerably different from the standard pronunciation. The absence of the /t/ sound affects the change in the structure of the consonants and the use of /eɪ/ instead of /ə/ affects the sounds produced by the vowel structure. The errors described could be due to interference from the speaker’s mother tongue or problems in pronouncing vowels and consonants in stressed and unstressed positions in English. Thus, this type of error can be classified into two types of errors, namely omission and substitution because it incorporates both of them by omitting the

consonants and replacing vowels. This implies that the production of consonants and reduction of vowels continue to be difficult for non-native teachers learning the language.

Datum 2.6



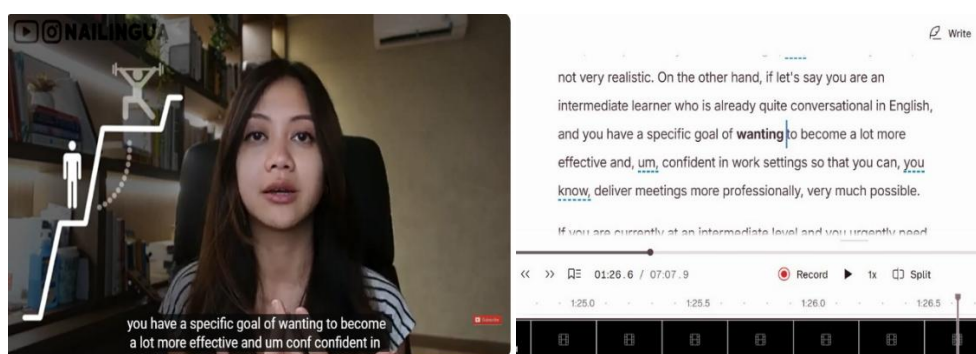
Datum 2.6 presents an example of a conversation which is based on an English learning video found on YouTube, where a tutor is teaching the students about something that pertains to study skills or study practices. In the course of teaching, the tutor makes use of the expression “your study routine and also estimating...” to provide information about the process of estimating something while learning. This interaction occurs in the online learning environment, where the use of English language is predominant in both teaching and learning processes. It is evident that the tutor is quite competent in conveying the message; however, some mistakes exist in the pronunciation of particular words.

In the statement above, the teacher used the word, “your study routine and also estimating...” For the word “estimating,” there is an example of mispronunciation in the word. The standard way of pronouncing the word “estimating” is “/’estimeɪtɪŋ/”. However, the teacher said it as “/’estimeɪnɪŋ/,” which was not a standard way of saying the word.

This deviation may occur because the consonant sound /t/, as the medial consonant, has been replaced by the sound /n/. According to Roach (2009), the sound /t/ is a voiceless alveolar plosive, a sound produced by placing the tip of the tongue on the alveolar ridge and then releasing air without any vibration of the vocal cords. Meanwhile, the /n/ sound is a voiced alveolar nasal, a sound produced with the same tongue position but where air is directed or expelled through the nasal cavity accompanied by vocal cord vibration.

The differences might result from the distinction in the manner of articulation, from a plosive to a nasal manner of articulation, as well as from the difference in voicing from voiceless to voiced. Despite the fact that the point of articulation (alveolar) is the same, the distinctions in the manner and nature of the sounds can cause substantial deviations from the pronunciation norms. The problem might be associated with the difficulties the speaker encounters in pronouncing certain consonants in spontaneous speech. Moreover, the phonological characteristics of the native language of the speaker can have an impact on the transformation of the sounds. Therefore, this error can be classified as a substitution, as it involves the replacement of one consonant sound with another consonant sound in the middle of a word. It indicates that differences in consonant articulation remain a challenge for non-native teachers.

Datum 2.7



The image shows a video recording interface. On the left, a woman is speaking. On the right, a text overlay displays a transcript of her speech. The transcript includes the following text:

not very realistic. On the other hand, if let's say you are an intermediate learner who is already quite conversational in English, and you have a specific goal of wanting to become a lot more effective and, um, confident in work settings so that you can, you know, deliver meetings more professionally, very much possible.

Below the transcript, there is a timeline with a play button and a record button. The timeline shows a duration of 01:26.6 / 07:07.9. The video player interface includes various controls like play, pause, and volume.

Datum 2.7 represents an interaction captured from a video lesson in English language learning on YouTube, where the teacher is conveying learning goals to the students. In this situation, the teacher clarifies that each student should set clear goals in their learning process. This interaction was conducted through an online learning session, where English language was used as the medium of instruction. The teacher seems to be quite proficient in presenting the information, but there are some errors in pronouncing some words.

In this statement, the teacher states “you have a specific goal of wanting...”. The problem here lies in the pronunciation of the word “wanting”. The correct pronunciation of the word “wanting” according to the standard rules of pronunciation is /'wɒntɪŋ/. She pronounced the word as /'wɒnɪŋ/.

Such deviation may result from the failure to pronounce the consonant /t/, which is a medial consonant in the word because it is found at the center of the word. Roach (2009) states that the omission of a sound may take place when there is the non-production of a sound that ought to be produced. The consonant /t/ in the word "wanting" is a voiceless alveolar plosive, which is produced through placing the tip of the tongue on the alveolar ridge.

Failure to pronounce the letter /t/ in the word would mean the simplification of the consonant cluster /nt/, making pronunciation of the word easier for the speaker. This process is called consonant cluster simplification. One of the consonants in the cluster is dropped to make it easier to pronounce.

The probable reasons behind this type of error include the difficulty of pronouncing consonant clusters in spontaneous English speech as well as the effects of the first language on the speaker's pronunciation habits. Therefore, this error can be

classified as an omission, as it involves the deletion of a consonant sound in the middle of a word. This finding indicates that the production of consonant clusters in English remains a challenge for non-native teachers in the process of using spoken English.

Based on this analysis the mispronunciation type of the 2 Indonesian online teachers from 10 short videos were various, the summary is presented in table 3 below.

Table 3. Summary of Analysis

No	Word (V/C)	Correct Pronunciation	Mispronunciation	Type	Place of Articulation	Position
1.	Picture (C)	/ˈpɪktʃər/	/ˈfɪktʃər/	Substitution	Bilabial - Labiodental	Initial
2.	Listen (C)	/ˈlɪsən/	/ˈlɪstən/	Addition	Alveolar	Medial
3.	Tree (C)	/tri:/	/dʒri:/	Substitution	Alveolar - Postalveolar	Initial
4.	Follow (V)	/ˈfɒləʊ/	/ˈfəloʊ/	Substitution	Back vowel – Front vowel	Medial
5.	Getting (C)	/ˈɡetɪŋ/	/ˈdʒetɪŋ/	Substitution	Velar - Postalveolar	Initial
6.	Answer (C)	/ˈænsər/	/ˈænswər/	Addition	Labio - Velar	Medial
7.	While (V)	/waɪl/	/wel/	Substitution	Diphthong - Monophthong	Medial
8.	Prepare (V)	/prɪˈpeə/	/prɪpər/	Substitution	Diphthong simplification	Medial
9.	Certain (C,V)	/ˈsɜ:tən/	/ˈsɜ:ɛn/	Omission & Substitution	Alveolar omission	Medial
10.	Estimate g (C)	/ˈestɪmeɪtɪŋ/	/ˈestɪmeɪnɪŋ/	Substitution	Alveolar - Alveolar	Medial
11.	Wanting (C)	/ˈwɒntɪŋ/	/ˈwɒnɪŋ/	Omission	Alveolar omission	Medial

B. DISCUSSION

1. Types of pronunciation error produce by online non-native teachers

Based on the results of the analysis in the findings section, this study identified three main types of pronunciation errors made by non-native English teachers conducting online lessons via the YouTube platform: substitution, addition, and omission. These three types of errors indicate deviations from standard English pronunciation involving both consonant and vowel sounds, and reflect difficulties with phonetic elements in the production of spoken English by non-native English teachers conducting online instruction. Pronunciation errors made by a speaker are analyzed according to phonetic standards using Peter Roach's theory (2009), which states that every speech sound in English can be analyzed through its articulatory characteristics, such as manner of articulation, place of articulation, and voicing. Thus, deviations in pronunciation can be analyzed through changes in these phonetic features.

Three types of errors that occurred, substitution was the most prevalent type found in the data of this study. This error was evident in various examples, both in consonant sounds and vowel sounds. In consonant sounds, substitution occurs in several words, namely picture (/p/ becomes /f/), tree (/t/ becomes /dʒ/), and getting (/g/ becomes /dʒ/); these errors indicate a sound change involving a difference in the place and manner of articulation of a sound. Additionally, regarding errors in vowel sounds, substitution occurs in the words "while" (/aɪ/ becomes /e/) and "prepare," where a diphthong has been simplified into a monophthong.

The findings indicate that substitution is the most common pronunciation error among Indonesian online English teachers. These results are consistent with the findings of Liu et al. (2025), who emphasize that pronunciation difficulties among non-

native English speakers generally involve segmental sounds due to differences between learners' first languages and the English sound system. However, unlike Liu et al. (2025), who focused on the application of Automatic Speech Recognition (ASR) for pronunciation learning and assessment, this study specifically examines pronunciation errors made by Indonesian online English teachers in authentic teaching videos. Therefore, this study expands on previous research by demonstrating that pronunciation errors are not limited to students but are also found among teachers who serve as pronunciation models.

The pattern above shows that speakers who make this error namely, non-native English online teachers tend to replace complex sounds with simpler ones or with sounds that are more frequently used in the phonological system of their first language; this is known as the phenomenon of substitution in pronunciation errors (Roach 2009). This finding is consistent with a study stating that phonological substitution is an adaptive strategy in speech production, whereby more complex sounds tend to be replaced with sounds that are easier for the speaker to produce (Alfaifi & Qasem, 20024).

The frequency of replacement found in this study suggests that pronunciation mistakes are systematic and connected to variations between the English and speakers' native tongues' sound systems rather than happening at random. The phenomena of L1 interference, also known as phonological transfer, which occurs when speakers project English sounds onto their preexisting sound system, is important in this context. As a result, the closest sound in terms of articulation is used to replace a sound that is unavailable in the native language (Odlin, 2003).

In addition to substitution, this study also identified another type of error, namely addition, which involves the insertion of sounds that should not be pronounced in standard English pronunciation. This error occurs in the words “listen” and “answer,” where speakers pronounce consonants that are essentially silent in English. This phenomenon occurs because of a speaker’s tendency to pronounce a word based on its spelling (orthography), rather than on the actual phonetic system of English. According to Petter Roach (2009), in English pronunciation, there are several types of letters that are not realized in speech sounds, so the pronunciation of a word does not always match its spelling. Therefore, a common challenge for non-native speakers is distinguishing between the written form and the phonetic form of a word.

The discrepancy between the spelling system and the sound system in English is one of the main factors causing this error. Rao (2018) explains that English has an irregular spelling system because there is no one-to-one correspondence between a letter and the sound it produces. In English, a single letter can produce several different sounds, and conversely, a single sound can be represented by a combination of different letters. This condition is the reason why English pronunciation can often be predicted solely based on the spelling.

Omission errors were also found in this research data, specifically in the words “wanting” and “certain.” In these cases, the speaker—an Indonesian non-native online English teacher—omitted the consonant sound /t/ located in the middle of the word, particularly when it occurred within a consonant cluster such as /nt/. The omission of this sound indicates a tendency to simplify the phonological structure of the word, known as consonant cluster simplification. Consonant clusters in English consist of a sequence of two or more consonant sounds pronounced in succession, which requires more complex articulation compared to single sounds (Roach 2009). Therefore, non-

native speakers often have difficulty maintaining all the consonant sounds in a single sequence of speech, so one of the sounds may be omitted to facilitate the articulation process.

This phenomenon generally occurs because the consonant clusters in English are considered more complex than those in the speaker's first language. In spontaneous speech, speakers tend to reduce the complexity of articulation by omitting one of the consonant sounds, thereby producing a form that is easier to pronounce. However, this simplified form of a word can cause changes in word form and may potentially affect the clarity of the utterance's meaning. This linguistic simplification phenomenon is known as simplification, which can be used to reduce language forms considered complex so that they are easier to produce and understand (Al-Safar 2021).

The results of this study also show that performance elements like speaking speed and communication context have an impact on pronunciation faults in addition to linguistic considerations. There is a greater chance of phonetic errors while pronouncing words since teachers in spontaneous learning procedures typically concentrate more on communicating the content of the subject than on pronunciation correctness. According to Artha & Bayu (2025), phonological mistakes, such as mispronouncing words correctly, can also interfere with communication fluency and result in miscommunication when speaking. Because speakers must speak immediately within a communicative context in spontaneous learning scenarios, there may be a higher chance of phonetic errors.

According to the study's overall findings, replacement is the most frequent mistake, followed by addition and deletion. According to this pattern, non-native speakers' main challenge is making target sounds that don't fit their phonological

system. As a result, they have to make adjustments by changing, adding, or removing specific sounds. Therefore, these three categories of mistakes show typical patterns in how non-native English speakers in Indonesia pronounce the language, especially while learning online. It is important to remember that professors act as role models for pupils' pronunciation. Students typically learn pronunciation by listening to and mimicking their teachers' pronunciations. Uzun (2022) explains that an English teacher's pronunciation ability is closely linked to the development of students' speaking skills. Consequently, pronunciation errors made by teachers impact the overall English learning process.

Findings from this study further indicate that pronunciation errors are likely to occur even in an online environment where non-native teachers facilitate learning. The same results have been found by Maxwell-Smith & Foley (2023). They indicated that bilingual individuals make different types of errors when pronouncing English sounds because of the effect of the phonological systems of their first languages. In addition, Ngo et al. (2024) point out that students learning English as a second language continue experiencing difficulties in keeping their pronunciation correct, especially in relation to those sounds which do not exist in the phonological systems of their native languages. Similarly, Sun (2023) points out that teaching pronunciation skills using technology increases pronunciation awareness despite the occurrence of errors during spontaneous pronunciation. Furthermore, Dai and Wu (2023) found out that technology-mediated pronunciation teaching allows sound errors detection among non-native speakers.

2. Realization of Speech Sound Errors in Terms of Articulatory Features

Based on the analysis presented in the findings section, the pronunciation errors made by non-native English teachers in online settings not only indicate a deviation

from standard English pronunciation but also reveal changes in certain articulatory features. Phonetic and phonological studies show that every sound in a language has articulatory characteristics that distinguish it from other sounds, such as place of articulation, manner of articulation, voicing, and the quality of the sound (Roach, 2009). Based on this, changes in one of these features can cause changes in the sound that affect the pronunciation of a word. Furthermore, phonological studies explain that every language possesses a distinct sound system, so differences in the sound systems between English and a speaker's first language can lead to deviations in word pronunciation (Yule, 2020). This finding aligns with the view of Ndububa (2025), who explains that non-native language learners often face difficulties in accurately producing English sounds; this is due to differences in the phonological systems between English and the learner's first language.

The most prominent change identified in this study is the change in the place of articulation that occurs in consonant sounds. This is demonstrated in the word "picture," where the sound /p/ changes to /f/. The sound /p/ is a voiceless bilabial plosive, while /f/ is a voiceless labiodental fricative. The shift in place of articulation represents a change in the sound from bilabial to labiodental. In addition to the example above, in the words "tree" (/t/ becoming /dʒ/) and "getting" (/g/ becoming /dʒ/), these changes indicate a shift in place of articulation from alveolar and velar to postalveolar. The changes occurring in these sounds indicate that speakers face difficulty in maintaining the target sounds in English, which are not fully present in the phonological system of their first language. This is explained in phonological studies (Odden, 2013). Jojua's (2025) research explains that in EFL learning, changes often occur in the substitution of consonant sounds due to the phonological system of the first language and articulatory limitations during the process of producing certain English sounds.

The changes observed in the findings pertain to the manner of articulation. These changes occur at the very beginning of sound production by altering the articulation in a way that differs from the standard pronunciation. An example of this can be found in the word “picture” (/p/), which is a plosive, changing to /f/, a fricative. A nearly identical change also occurs in the words “getting” and “tree,” where the sounds /g/ and /t/—which are plosives—change to /dʒ/, an affricate. Roach (2009) states that changes in the manner of articulation can significantly affect sound quality because the mechanisms involved in the airflow used are different. Anjani et al. (2023) state that EFL learners often experience difficulties when producing English consonant sounds that involve differences in manner of articulation, particularly those occurring in the fricative, affricate, and consonant sound systems that do not exist in the sound system of their first language.

In the findings above, changes were also observed in the voicing feature. This change occurred in the word “estimating,” where the /t/ sound changed to /n/. Both sounds share the same place of articulation, which is alveolar. The /t/ sound is a voiceless alveolar plosive, while /n/ is a voiced alveolar nasal (Roach, 2009). These changes suggest that speakers experience alterations in vocal cord vibration and airflow mechanisms. Changes in voicing features can also affect the phonological identity of voiced and voiceless sounds, which possess distinct acoustic characteristics, as explained in phonological studies (Yule, 2020). Differences in the sound systems between English and the speaker’s first language can cause difficulties in producing English sounds more accurately, leading to a variety of sound changes in EFL speakers’ pronunciation (Gharamah et al., 2024).

This study also revealed changes in vowel quality and the simplification of diphthongs. The diphthong /a—/ in the word “while” transformed to the monophthong

/e/. A diphthong, according to Roach (2009), is a combination of two vowel characteristics in a single syllable that requires a change in tongue position during articulation. But in this study, diphthongs were reduced to monophthongs, which simplified the articulatory actions. This shift occurred in the word "prepare," where the monophthong /a/ replaced the diphthong /e." The vowel quality of the word "follow" changed from back-rounded to unrounded. These changes suggest that the English language system remains a challenge for non-native speakers, particularly when the vowel sounds do not correspond to those in their first language. Phonological studies indicate that differences in vowel systems across languages are often the primary cause of pronunciation errors among foreign language learners (Fromkin et al., 2018).

In addition to the changes in single consonant sounds described above, this study also found instances of consonant cluster simplification. In the words "wanting" and "certain," the /t/ sound in the medial position was omitted by the speakers. Consonant clusters in English require more complex articulatory coordination because they involve two or more consonant sounds pronounced in sequence. Non-native speakers often face difficulty in maintaining all consonant sounds within a single speech sequence. Non-native speakers also frequently struggle to retain all consonant sounds within a single speech sequence; consequently, one of the sounds tends to be omitted. This process of simplifying consonant clusters is known as phonological simplification, which is a tendency for a speaker to reduce sound structures considered complex to make them easier to pronounce (Odden, 2013).

Therefore, based on the results of this study, it can be observed that the production of English sounds by non-native speakers is affected by the complicated nature of the English sound system. This is similar to what is stated in the research conducted by Maxwell-Smith and Foley (2023), who found out that bilingual speakers are still faced

with problems in producing English sounds as influenced by the sound system of their native language. In addition, according to Ngo et al. (2024), EFL learners are still prone to making several pronunciation mistakes because of the existence of consonant and vowel sounds that are not present in their native language. The findings in the study conducted by Wei et al. (2022) also prove that non-native speakers struggle in sustaining accurate pronunciation especially for segmental sounds like consonants and vowels. As asserted by Farrús (2023), pronunciation mistakes in non-native speakers are not random at all; rather, they are influenced by articulation difficulties in phonological systems.

The participants' first-language background is a key factor in understanding the pronunciation errors identified in this study. All teachers selected as study participants were non-native English speakers from Indonesia whose first language is Indonesian. Differences between the phonological systems of English and Indonesian may contribute to the occurrence of pronunciation errors because speakers tend to transfer phonological patterns from their first language when producing sounds in a second language or foreign language (Odlin, 1989). Some English sounds, such as the dental fricatives (/θ/ and /ð/), certain diphthongs, and consonant clusters, have no direct equivalents in Indonesian. As a result, Indonesian speakers may substitute unfamiliar sounds with more familiar ones from their native language, leading to pronunciation errors such as substitution, addition, or omission. Therefore, the participants' first-language background is considered a significant factor that may influence the occurrence of the pronunciation errors analyzed in this study.

The influence of the participants' first language observed in this study is consistent with Roach's (2009) explanation, which states that differences between sound systems can lead to systematic pronunciation errors. Since all participants in this

study were Indonesian speakers whose first language was Indonesian, certain English sounds that do not exist in Indonesian tended to be substituted with more familiar sounds. This finding also supports the results of a study by Liu et al. (2025), which reported that learners' linguistic backgrounds play an important role in shaping their pronunciation performance.

CHAPTER V

CONCLUSION AND SUGGESTION

This chapter presents the conclusions drawn from the findings and discussion in the previous chapter. The summary in this chapter focuses on the types of errors made and their characteristics. In addition, this chapter offers several suggestions for further research.

A. CONCLUSION

Based on the results and discussion, this study discovered that non-native Indonesian English teachers in online learning made a number of pronunciation mistakes. The three primary categories of pronunciation problems identified in this study are addition, omission, and substitution. Substitution was the most prevalent of the data which include consonant swaps, consonant cluster omissions, superfluous sound additions, and diphthong simplifications, that affect both vowel and consonant sounds. It shows that the distinctions between the phonological systems of English and the speakers' native tongues have an impact on the pronunciation challenges faced by non-native Indonesian English teachers.

Additionally, this study discovered that these pronunciation errors reflect a number of changes in articulatory parameters, including vowel quality, consonant cluster simplification, diphthong simplification, voicing, site of articulation, and mode of articulation. These alterations show that the speaker's articulatory modifications during the production of English sounds are connected to the systematic pronunciation problems. The findings of this study also show that English sounds that are considered

difficult or unfamiliar tend to be simplified by non-native speakers to make them easier to produce in the communication process.

. The pronunciation errors identified in this study can be understood as a form of articulatory adaptation employed by speakers to facilitate the speech production process in oral communication.

B. SUGGESTION

Several recommendations for English language instruction methods and additional research can be made in light of the study's conclusions and limitations. First, since this study indicate that the pronunciation errors do not occur randomly but are related to changes in specific articulatory features, non-native English teachers are required to focus more on pronunciation correctness by assessing their own speaking performances or instructional films. Second, because the occurring errors indicate the influence of the first language's sound system on the production of English sounds, it is recommended that future researches investigate pronunciation problems in greater detail using more data and people from different educational backgrounds and geographical areas. In order to examine additional factors impacting pronunciation errors, such as the impact of the first language, speaking habits, and teaching experience, future study could potentially integrate phonological analysis with interviews or observations. Furthermore, it is suggested that future researchers employ phonological theories other than Roach's to provide a more comprehensive explanation for articulatory characteristics and pronunciation problems by using Automatic Speech Recognition (ASR) or other digital technologies.

Although this study has discovered a variety of articulatory characteristics and pronunciation problems displayed by non-native English teachers in Indonesia, it still

has a number of limitations. First, there are no direct observations or interviews with the research subjects in this study; instead, it is entirely focused on data analysis. Therefore, it is not yet possible to provide a more thorough explanation for other elements like the background of pronunciation learning, advancements in phonological competence, or the habits of English speakers. Second, only non-native English teachers from Indonesia provided the data for this study. As a result, the results do not yet adequately describe the particular pronunciation mistakes made by non-native teachers according to their first-language backgrounds. Third, this study only uses Roach's (2009) theory on phonetics and phonology as the primary foundation for analyzing pronunciation and articulatory features. Although this theory is relevant to this study, relying on a single theory may limit the perspective during the analysis process.

REFERENCES

- Creswell, J. W. (2012). *Research design: Pendekatan kualitatif, kuantitatif, dan mixed*. Pustaka Pelajar.
- Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, 36(5–6), 861–884.
<https://doi.org/10.1080/09588221.2021.1952272>
- Dillon, T., & Wells, D. (2023). Effects of pronunciation training using automatic speech recognition on pronunciation accuracy of Korean English language learners. *English Teaching*, 78(1), 3–23. <https://doi.org/10.15858/engtea.78.1.202303.3>
- Farrús, M. (2023). Automatic speech recognition in L2 learning: A review based on PRISMA methodology. *Languages*, 8(4), Article 242. <https://doi.org/10.3390/languages8040242>
- Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, 12(1), Article 2466288. <https://doi.org/10.1080/2331186X.2025.2466288>
- Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge.
- Maxwell-Smith, Z., & Foley, B. (2023, May). Automated speech recognition of Indonesian-English language lessons on YouTube using transfer learning. In *Proceedings of the Second Workshop on NLP Applications to Field Linguistics* (pp. 1–16).
<https://doi.org/10.18653/v1/2023.fieldmatters-1.1>

- Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/S0958344023000113>
- Odden, D. (2013). *Introducing phonology* (2nd ed.). Cambridge University Press.
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.
- Spring, R., & Tabuchi, R. (2021). Assessing the practicality of using an automatic speech recognition tool to teach English pronunciation online. *Journal of English Teaching through Movies and Media*, 22(2), 93–104. <https://doi.org/10.16875/stem.2021.22.2.93>
- Sugiyono, S. (2007). *Metode penelitian kualitatif, kuantitatif, dan R&D*. Alfabeta.
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, 14, Article 1210187. <https://doi.org/10.3389/fpsyg.2023.1210187>
- Wei, X., Cucchiarini, C., van Hout, R., & Strik, H. (2022). Automatic speech recognition and pronunciation error detection of Dutch non-native speech: Cumulating speech resources in a pluricentric language. *Speech Communication*, 144, 1–9. <https://doi.org/10.1016/j.specom.2022.08.004>
- Yule, G. (2020). *The study of language* (7th ed.). Cambridge University Press.
- Alfaifi, A., & Qasem, F. (2024). Phonological substitution patterns in Yemeni Ibbi Arabic child speech: a markedness and natural phonology perspective. *Cogent Arts & Humanities*, 11(1). <https://doi.org/10.1080/23311983.2024.2360795>

- Odlin, T. (2003). Cross-linguistic influence. *The handbook of second language acquisition*, 436-486. DOI:10.1002/9780470756492
- Rao, V. C. (2018). English spelling and pronunciation: a brief study. *J. Res. Sch. Profess. Eng. Lang. Teach*, 2, 1-10.
- Al-Safar, O. W. S., & Al-Halawachy, A. A. I. Simplification Strategy and the Comprehension of Some Parts of Speech: Need or don't Need to be Simplified?.
- Artha, D. J., & Bayu, A. (2025). Analyzing EFL Learners Phonological Errors Using Speech Recognition Technology Soundtype AI. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 13(1), 2745-2757. <https://doi.org/10.24256/ideas.v13i1.6845>
- Uzun, T. (2022). PRESERVICE ENGLISH TEACHERS' ATTITUDES AND BELIEFS ABOUT LEARNING AND TEACHING PRONUNCIATION. *Dil Dergisi*, 173(2), 22-42.
- Ndububa, C. L. (2025). Challenges in teaching english speech sounds to non-native speakers using british dictionaries. *International journal of academic pedagogical research (IJAPR)*, 9(4), 74-86.
- Joju, N. (2025). Consonant Articulation and Lexical Stress in Georgian EFL Learners: Phonological Patterns and Pedagogical Implications. *MULTILINGUAL EDUCATION*, 79, 104.
- Anjani, R. P. W., Wahyuni, H. I. W., Nugraheni, M. P. P., & Bram, B. B. B. (2023). Problematic English Consonant Sounds for Indonesian EFL Learners. *Journey: Journal of English Language and Pedagogy*, 6(3), 592-604.

Awadh, F. M. A. M., Putro, N. H. P. S., Pohan, A. E., & Alsamiri, Y. A. (2024). Improving English pronunciation through phonetics instruction in Yemeni EFL classrooms. *JOLLT Journal of Languages and Language Teaching*, 12(2), 930-940.

Gilakjani, A. P. The Role of EFL Teachers in Improving EFL Learners' English Pronunciation

CURRICULUM VITAE



Refi Ananda was born in Lumajang on June, 28 2003. She graduated from MIN 01 Lumajang in 2016, then from SMP Tri Sakti in 2019, and from SMA Manarul Qur'an Boarding School in 2022. She participated in student council (OSIS) activities during junior high and senior high school. Additionally, she served as a member, and board members of the boarding school Manarul Qur'an. After graduating from

SMA Manarul Qur'an Boarding School, she continued her studies at UIN Maulana Malik Ibrahim Malang in 2022 in the Faculty of Humanities, majoring in English Literature. During her undergraduate studies, she actively participated in all courses and completed her final thesis to achieve a Bachelor's degree in English Literature (S.S.).

APPENDIX

Datum	Utterance	Word	Standard Pronunciation	Produced Pronunciation	Type of Error
Datum 1.1	“Let’s look at the next picture.”	Picture	/ˈpɪktʃər/	/ˈfɪktʃər/	Substitution
Datum 1.2	“It’s important to listen carefully.”	Listen	/ˈlɪsən/	/ˈlɪstən/	Addition
Datum 1.3	“This is a tree.”	Tree	/tri:/	/dʒri:/	Substitution
Datum 1.4	“But I do believe that I think getting all...”	Getting	/ˈgetɪŋ/	/ˈdʒetɪŋ/	Substitution
Datum 2.1	“Please answer the question carefully.”	Answer	/ˈɑːnsər/	/ˈɑːn.swər/	Addition
Datum 2.2	“You should prepare your study plan.”	Prepare	/prɪˈpeər/	/prɪˈpər/	Substitution
Datum 2.3	“You have to follow the instruction.”	Follow	/ˈfɒləʊ/	/ˈfæləʊ/	Substitution
Datum 2.4	“You can do it while studying.”	While	/waɪl/	/wel/	Substitution
Datum 2.5	“When you’re feeling a certain way.”	Certain	/ˈsɜːtən/	/ˈsɜːeɪn/	Omission & Substitution
Datum 2.6	“Your study routine and also estimating...”	Estimaning	/ˈestɪmeɪtɪŋ/	/ˈestɪmeɪnɪŋ/	Substitution
Datum 2.7	“You have a specific goal of wanting...”	Wanting	/ˈwɒntɪŋ/	/ˈwɒnɪŋ/	Omission

Word	Sound Changes	Place of Articulation	Manner of Articulation	Voicing	Articulatory Feature
Picture	/p/ → /f/	bilabial → labiodental	plosive → fricative	voiceless → voiceless	place & manner change
Listen	/s/ → /st/	alveolar	fricative → fricative + plosive	voiceless	addition
Tree	/t/ → /dʒ/	alveolar → postalveolar	plosive → affricate	voiceless → voiced	place, manner & voicing change
Getting	/g/ → /dʒ/	velar → postalveolar	plosive → affricate	voiced → voiced	place & manner change
Answer	addition of /w/	labio-velar	approximant	voiced	addition
Prepare	/eə/ → /a/	central vowel simplification	diphthong → monophthong	voiced	diphthong simplification
Follow	/ɒ/ → /æ/	back vowel → front vowel	rounded → unrounded	voiced	vowel quality
While	/aɪ/ → /e/	diphthong → monophthong	vowel simplification	voiced	diphthong simplification
Certain	omission of /t/	alveolar omission	plosive deletion	voiceless omission	consonant cluster simplification
Estimating	/t/ → /n/	alveolar → alveolar	plosive → nasal	voiceless → voiced	manner & voicing change
Wanting	omission of /t/	alveolar omission	plosive deletion	voiceless omission	consonant cluster simplification

