CHAPTER II
REVIEW OF RELATED LITERATURE

In this chapter, the researcher discusses about social media for learning pronunciation, TikTok, pronunciation, artificial intelligence in English learning, and previous study.

A. Social Media for Learning Pronunciation

Cebrera et. al. (2017) that cited in Allen et al. (2023) defines social media as digital and electronic platforms that can support communication and let the users to create and share contents. Since its appearance, social media has grown up and gained its popularity rapidly. There are many people who use social media. In Indonesia, over 76% women and 75% men above five years old use social media (Direktorat Statistik Kesejahteraan Rakyat, 2023). It proves that most people in Indonesia use social media.

Usually, people use social media for many reasons. One of the reasons is learning English, especially pronunciation. The contents about pronunciation get a positive perception from English learners. Safitri, et. al. (2022) stated that social media facilitate English contents to help students practicing their English skills and widening their vocabulary and pronunciation. Moreover, Utami (2018) stated that social media is very useful for improving language skills, learning motivation, and self-confidence. The language skills that dominantly enhanced were pronunciation and speaking.

The positive perception of social media given by the users is not unreasonable. By utilising social media, people can get more knowledge about English, especially pronunciation. The study conducted by Coronel (2022) showed that the implementation of social media in the classroom can

improve students' pronunciation. In his research, he applied social media which are Facebook, WhatsApp, Instagram, and TikTok as a lecturing media. The result showed that all students in experimental group have grades equal to or more than 10.5 out of 15 after the treatment. However, in control group, there are some studets who have a score below 10.5 out of 15. Besides, Ilyas and Putri (2020) also proved that utilising social media in teaching speaking can improve students' speaking skill, including vocabulary, grammar, and pronunciation. The average score for pronunciation was increased from 49.46 in pre-test becoming 61.63 in post-test. These studies showed that social media can be an option to learn English pronunciation.

B. TikTok

TikTok is a video-sharing social media platform from China (Omar & Dequan, 2020). It is maintained by ByteDance and started in September 2016. This media app is accessible in over one hundred fifty marketplaces and in seventy five languages (Jaffar et al., 2019). This platform is very popular in the world. Chen and Kang (2023) stated that this application attracts active users and become the most downloaded application for two years. Besides, the data from Data Reportal (2023) that is cited in Pèrez-Sabater, et. al. (2023) showed that, in April 2023, TikTok had 834.3 million monthly users all over the world.

Based on the data from Wu (2020) that is cited in Chen and Kang (2023), TikTok users are dominated by young generation under thirty years old and spend for about fifty two minutes in a day. Users can create and share

short videos with the duration started from fifteen seconds to ten minutes. Besides, users also can give comments or likes to other videos.

The content of the videos uploaded on TikTok are various. Kåhlman (2023) stated that the videos on TikTok have many genres such as tricks, jokes, dance, and other entertainment including educational contents. The result of the survey from Study.com that is cited in Mahat and Boulghalaghe (2023) showed that the contents that have the highest number of likes on TikTok are English Videos. They also stated that spelling and pronunciation contents got almost one million likes.

With many contents about education, TikTok is considered to be a new way to learn something. Opas (2023) stated that TikTok can be an interesting option to help teaching and learning because of its length and features that can improve students' learning ability. In terms of pronunciation contents, Fitria (2023) stated that short video is very suitable for this type of contents and TikTok is very attractive as a platform to learn English.

Based on students' perception, TikTok is a good platform to learn pronunciation. It was already mentioned by some researchers. The research conducted by Mahat and Boulghalaghe (2023) showed that learning English pronunciation from TikTok is more enjoyable and effective than learning it from textbook or classrooms. Besides, Bahri et al. (2022) also mentioned that the English contents on TikTok are about vocabulary, grammar, pronunciation, and common mistakes. They stated some reasons why TikTok is good for learning English are because the video is short, the explanation is clear, and the caption can give a positive impact in learning process.

In terms of teaching and learning, TikTok is also impactful to improve students' pronunciation skill. Some studies showed that TikTok has a significant effect on students' pronunciation (Nasution & Nurlaili, 2023; Nuari, 2022). In line with this statement, the research conducted by Rahmawati, et. al. (2023) showed that using TikTok in a classroom can improve students' pronunciation until thirty-one point six percent. Khikmah, et. al. (2024) also stated that using TikTok as learning media can reduce students' pronunciation mistake. It shows that TikTok is useful in a teaching and learning process.

C. Pronunciation

Pronunciation is one of the crucial aspects in communication. Pennington and Rogerson-Revell (2019) state that pronunciation is the significant aspect of oral languages since an idea is necessary to be spoken in order to be apprehended and to be considered as messages that can be talked to others. Mistakes in pronunciation may lead to some misunderstandings. If someone's pronunciation is difficult to understand, although their grammar and vocabulary are great, their communication will be unsuccessful (Yoshida, 2016). In line with this statement, Sardegna and Jarosz (2023) stated that the mistakes in pronunciation may lead to misunderstanding and communication failure that can cause serious effect. It can be worse than making mistakes in Grammar or vocabularies (Jarosz, 2019).

Since the English pronunciation is important, people start to learn this aspect of language. Low (2015) stated that, nowadays, the awareness of the importance of pronunciation in learning English as an international language

is growing. With great pronunciation ability, someone will also have more self-confidence when they talk in front of many people (Kobilova, 2022; Low, 2015).

Kráľová, et. al. (2021) define pronunciation as the sound production that is utilised to communicate meaning. Pronunciation can be divided into two categories which are suprasegmental and segmental features (Dawson & Phelan, 2016).

Suprasegmental is an aspect of pronunciation that affect more than just a single sound such as intonation, stress, and rhythm (Yoshida, 2016). Stress is a more forceful production of a particular syllable because of parameters of increased length, loudness, higher pitch, and syllables that possess a full vowel quality (Kang et al., 2018). Rhythm is the result of timing regularity of similar or like events and give rise to the notion of isochrony (Kang et al., 2018). Intonation is the way of someone's voices fall and rise in pitch throughout a whole sentence or utterance (McMahon, 2020).

In the opposite, segmental features are the individual phonemes of a language which are consonants and vowels (Yoshida, 2016). A vowel is a sound where the airstream raises from the lungs and smoothly passes the vocal tract without anything blocks it. Moreover, consonants are speech sounds that involve an obstruction of the airstream when it passes the vocal tract (Carley & Mees, 2024).

McMahon (2020) stated that vowels are classified into two types which are monophthongs and diphthongs. Monophthongs are vowels that are its quality remains consistent since the beginning of its production until the end.

Based on its length, monophthongs can be divided into tense vowels and lax vowels. There are five tense vowels of monophthongs which are /iː/, /uː/, /ɑː/, /ɔː/, and /ɜː/. In some references such as Carley and Mees (2020); Fromkin, et. al. (2018); and Yule (2020), they do not add /ː/. So, the symbols become /i/, /u/, /ɑ/, /ɔ/, and /ɜ/. In Cameron (2018), a difference also appears. If there is unstressed syllables of "er" or "or", it is not written /ər/ but /ɚ/. Besides, there is no /ɜ/. This is changed into /ɝ/. /ɝ/ sounds like /ɜr/. This theory is also mentioned in Yoshida (2016). In Yoshida (2016), /ər/ can bi written /ɚ/ and /ɜr/ can be written /ɝ/. While there are some variations of symbols in those vowels, the remained monopthongs have the same symbols. There are six lax vowels of monophthongs which are /ɪ/, /ʊ/, /ə/, /ɛ/, /ʌ/, and /æ/. While monophthongs are consistent in its quality, diphthongs are the opposite. Its quality changes during its production. All diphthongs are long vowels. There are five diphthongs in General American which are /oʊ/, /aʊ/, /eɪ/, /ɔɪ/, and /aɪ/.

Vowel can also be classified based on the tongue and lip position (Brown, 2014). The tongue position is divided into two main categories which are vertically and horizontally (McMahon, 2020). Vertically, the position of the tongue is high (close to the roof of the mouth), or low (far from the roof of the mouth with the jaw open), or somewhere in between (mid). Vowels that are categorised as close (high) are /iː/, /ɪ/, /ʊ/, and /uː/. In mid vowel categories, there are four sounds which are /ɛ/, /ɜː/, /ə/, and /ɔː/. Open vowels also have three sounds which are /æ/, /ʌ/, and /ɑː/. Horizontally, the position is far forward in the mouth (front), or at the back (back), or

somewhere in between (central). The sounds that are categorised as front vowels are /ɪ/, /ɛ/, /æ/, and /iː/. The sounds that are categorised as back vowels are /ɑː/, /ɔː/, /ʊ/, and /uː/. The last, there are three central vowels which are /ɜː/, /ə/, and /ʌ/. Moreover, based on the lips position, the vowels divided into rounded or unrounded. The rounded vowels are /ʊ/, /ɔː/, and /uː/ and the rest vowels are unrounded (Dawson & Phelan, 2016). However, in other theory, it is mentioned that rounded vowels are /ʊ/, /ɔː/, /uː/, and /ɑ/ (Brown, 2014)

As well as vowels can be categorised based on three conditions which are its quality, tongue position, and lip rounding, consonants also can be categorised based on three conditions. However, the conditions are different with vowels. In consonants, the three factors that must be considered are voicing, place of articulation, and manner of articulation (Carley & Mees, 2024; Kang et al., 2018; Yoshida, 2016).

Voicing refers to the present of the vibration in the vocal cords (Kang et al., 2018). Based on this term, consonants are divided into two categories which are voiced and voiceless. Voiced means the sounds produced when the vocal cords are close together and tightened. The consonants that categorised as voiced are /b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/, /ʤ/, /m/, /n/, /ŋ/, /j/, /w/, /r/, and /l/. On the other hand, voiceless is produced when the vocal cords are open and relaxed then air passes freely through the open glottish with no vibration. The consonants that categorised as voiceless are /p/, /t/, /k/, /f/, /ɵ/, /s/, /ʃ/, /h/, and /ʧ/.

Second, place of articulation is a place where an obstruction is made in the vocal tract (Carley & Mees, 2020). The place of articulation is divided into seven categories which are bilabial, labiodental, dental, alveolar, palatal, velar, and glottal. the First, bilabial means both lips is touching or almost touching. There are four sounds in this group, which are /p/, /b/, /m/, and /w/. Second, labiodental means the upper teeth is touching the lower lip softly. There are two sounds in this group, which are /f/ and /v/. Third, dental means the tongue tip is touching the upper front teeth or between the front teeth. There are two sounds in this group which are /ɵ/ and /ð/. Fourth, alveolar means the tongue tip is touching the alveolar ridge. There are six sounds in this group which are /t/, /d/, /s/, /z/, /n/, and /l/. Fifth, palatal means the blade of the tongue is touching or almost touching the hard palate. There are six sounds in this group which are /ʃ/, /ʒ/, /ʧ/, /ʤ/, /r/, and /y/. Sixth, velar means the back of the tongue is touching the soft palate. There are three sounds in this group which are /k/, /g/, and /ŋ. Seventh, glottal means there is a friction in the glottis. In this group, there is only one sound which is /h/.

Last, manner of articulation is the term used to describe the type of obstruction existed in pronouncing a consonant (Carley & Mees, 2024). There are five manners of articulation which are plosive, fricative, affricate, nasal, and approximant. First, plosive is produced by blocking the air flow completely, and then released. Consonants that are categorised as plosives are /p/, /b/, /t/, /d/, /k/, and /g/ (Carley & Mees, 2024). Second, fricative is sounds that are produced via close approximation of two articulators so that airstream is partially obstructed and turbulence results. Consonants that are categorised

as fricatives are /f/, /v, /θ/, /ð/, /s/, /z/, /ʃ/, /ʒ/, and /h/. Third, affricate is a combination of a stop followed by a fricative, an explosion with a slow release. Consonants that are categorised as affricates are /ʧ/ and /ʤ/. Fourth, nasals are the sounds that produced by the tongue or lips block off the vocal tract so air can't go out through the mouth. Consonants that are categorised as nasals are /m/, /n/, and /ŋ/. Fifth, approximants are sounds that produced when a narrowing is formed in the vocal tract, but one not narrow enough to cause a hissing noise as in the case of a fricative. Consonants that are categorised as approximants are /j/, /w/, /r/, and /l/.

Some consonants have different pronunciation based on their position in a word. For example, in consonants, there are two groups which are obstruents and sonorants (Carley & Mees, 2020). Obstruents are voiced and voiceless pairs while the others are sonorants including vowels. Voiced and voiceless pairs are plosive (/p b/, /t d/, and /k ɡ/), affricates (/ʧ ʤ/), and fricatives (/f v/, /θ ð/, /s z/, and /ʃ ʒ/). Voiceless obstruents lessen sonorants that are in front of them in the same syllable. For example help /hɛlp/, bench /bɛnʧ/, and bank /bæŋk/. While voiceless obstruents can shorten sonorant, voiced obstruents are not. However, they have a different articulation based on their position. When voiced obstruents are between sonorants, they are typically fully voiced. For example rabbit /ˈræbət/, colder /ˈkoʊldər/, and cargo /ˈkɑrgoʊ/. Moreover, voiced obstruents are typically partially or fully devoiced when they are preceded by a pause or a voiceless consonant or when a pause or voiceless consonant follows. For example this book /ˈðɪs ˈbʊk/, dog /ˈdɑg/, and grab some /ˈgræb ˈsəm/.

Another aspect that must be considered in pronouncing consonant is aspiration. Aspiration is a brief period of voicelessness between the release of the plosive and the beginning of voicing for the next sounds (Carley & Mees, 2020). It occurs when voiceless plosives which are /p/, /t/, and /k/ are at the beginning of a stressed syllable. When /p/, /t/, and /k/ are not at the beginning of a stressed syllable, at the end of syllables, or when preceded by /s/ at the beginning of a syllable, the aspiration is weak or unaspirated. For example perform /pərˈform/, goat /goʊt/, and score /skor/. Moreover, when approximants (/l r w j/) come after /p t k/, the aspiration happens during the articulation of approximants that makes the approximant partially or fully devoiced and cause turbulence and fricative noise at the place of articulation of the approximants. For example play /plei/, cream /krim/, and cue /kju/. However, the approximants are not devoiced or become fricative if there is /s/ at the beginning of a cluster. For example sprint /sprɪnt/ and scream /skrim/.

The /t/ sound has additional allophone that is glottal stop (ʔ) (Yoshida, 2016). To produce this sound, the vocal cords close tightly, air builds up behind them, and then they open quickly. It's like the beginning of a small cough, or the middle sound when we say huh-uh to mean "no." /t/ can be pronounced as a glottal stop when the syllable before it is stressed and the syllable after it is unstressed or when the syllable after it is /ən/ or syllabic /n/.

Not only glottal stop, can /t/ also be pronounced as flap or tap (Carley & Mees, 2020; Fromkin et al., 2017; Yoshida, 2016). This happens when /t/ is between two vowels such as city, vitamin, and critic. In those words, /t/ is flapped which is pronounced voiced and indistinguishable with /d/. It is

written with symbol /ɾ/. Besides /t/, /d/ is also flapped. Flapped /d/ is a quick, soft /d/ or /r/ sound. The condition and the symbol are same as flapped /t/.

In addition to plosive, there is a term called stop sequence. Carley and Mees (2020) explained that there are several conditions that can make stop sequence occurs. First, when two identical plosives occur in sequence, usually, the first plosive is not released and it sounds like there is a long plosive e consisting of an approach stage followed by a long hold stage and then a release stage such as ripe pear, get two, and black car. Second, the plosives that have the same place of articulation but differ in voicing is also unreleased when they are in a sequence such as stop by, hot dinner, and black gown. Third, when plosives occur in sequence at different places of articulation, the closure for the second plosive is made before the closure for the first plosive is released. Consequently, the release of the first plosive is inaudible because a closure has already been made further forward in the mouth or because a closure further back in the mouth holds back the compressed air. For example bad place and grab two. In opposite, when affricate occurs as the first in a sequence of stops, it is always released. The fricative release stage of affricates is always present, for example rich cheese and much pain.

D. ELSA Speak for Pronunciation Analysis

ELSA Speak is a mobile application for pronunciation learning that using artificial intelligence and speech recognition (Fitria, 2021). The app has capabilities including analysing speech, creating lesson sets, having a dictionary feature, finding relevant English lessons, providing a coach or

tutor, providing interactive and fun game features, and having the skill of measuring students' abilities (Arbi, 2024). It is developed by Vu Van in 2015 and the company is located in San Fransisco, California (Marlinda and Huda, 2024). It can be downloaded on Google Play or App Store. Nushi and Sadeghi (2021) reported that this platform is already downloaded from Google Play and Apple Store for more than fifteen millions.

The study conducted by Widyasari and Maghfiroh (2023) showed the benefits of ELSA Speak application. It is mentioned that ELSA Speak can be downloaded for free, detects pronunciation with 90% accuracy, provides feedback for users, have one thousand and two hundred lessons with more than sixty topics, and offers an interactive dictionary. On another research that conducted by Luu et al. (2021), it is stated that the accuracy of pronunciation detection rate reaches 95% (cited in Marlinda and Huda, 2024). It proves that ELSA Speak is an acceptable platform for analysing pronunciation.

The pronunciation analysis feature on ELSA Speak gives some benefits for its users. Arbi (2024) stated that ELSA Speak facilitate specific feedback related to pronunciation and intonation based on the result of its speech analysis feature. Moreover, Nushi and Sadeghi (2021) mentioned that the accurate speech recognition technology from ELSA Speak combined by artificial intelligence can help users recognising the slightest differences between their pronunciation and native models.

E. Previous Study

There are some previous studies analyse the pronunciation of someone. The first study was conducted by Rios-Urrego et.al. in 2023, entitled 'Automatic Pronunciation Assessment of Non-native English Based on Phonological Analysis.' The goal of this research is analysing a non-native English speakers' pronunciation ability. The data were taken from pronunciation test results of the respondents. The results of the test were checked in terms of correctness and incorrectness. The students' skills and mistakes English long and short vowel pronunciation were analysed by the researcher. The researcher also calculated the students' skills and errors. The results of the study showed that students got 840% total score in short vowel and 804% total score in long vowel. From the long vowel pronunciation test, it was revealed that there are three students got very good scores and seven students got good scores. Moreover, in the short vowel pronunciation test, there are eight students who got very good scores and two students got good scores.

The second study was conducted by Mahjabeen, et. al. in 2024, entitled 'An Acoustic Analysis of Diphthongs in Pakistani and British English. The purposes of the study are to find out how the Basic English pronunciation skill of English Study Program students is, what their difficulties in segmental and supra-segmental aspect of pronunciation are, and the categories of Basic English pronunciation that they can pronounce them well. This study used quantitative method. The sample was taken from twenty-six students at the age of eighteen to twenty-five years old that took English study program in

State University of Timor-Indonesia. The instruments of the study are list of one hundred Basic English vocabularies and a tape recorder. The students were asked to read and speak some words while the researcher recorded it. The data were analysed by using codification, categorization, tabulation, and discussion. The score was based on segmental and supra-segmental aspects of pronunciation. The scale of the score started from one as the lowest score (bad), two (less), three (enough), and four as the highest score (accurate). Based on the results, students only got 'enough' as the highest score in supra-segmental aspect, which are 16 students (61.5%) of 'Stress' element, 17 students (65.3%) of 'Intonation' element, 12 students (46.1%) in 'Rhythm', and 13 students (50.0%) of 'Voice Quality.' In segmental aspect, the highest score also in 'enough' level, which are 16 students (61.5%) got this score in 'Gestures.' Meanwhile, the element that got the lowest score was 'rhythm'. Most students (53.8% = 14 students) got the 'less' score. However, no one got 'bad' score.

The third study was conducted by Yusuf, et. al. (2024), entitled 'The Sounds of Indonesian English: Acoustic Phonetic Analysis of the Monophthong Vowels across Genders.' This study analysed the monophthong vowels sounds production of Acehnese English for foreign language learners. This study used quantitative method. The informants were the fourth semester undergraduate students at a private college in Aceh. The data were measured using tools named PRAAT 5.3.5.3. This study revealed that females pronounced some vowel pairs in similar, which are /iː/-/ɪ/, /ɛ/-/æ/, /uː/-/ʊ/, /ʌ/-/ɑː/, and /ɑː/-/ɒ/. In the opposite, they can differentiate the

sounds of some pairs, which are /ɜ:/-/ʌ/ and /ɔ:/-/ɒ/. However, they cannot distinguish long and short vowel pairs. Meanwhile, males can differentiate these pairs: /i:/-/ɪ/, /ɜ:/-/ʌ/, /ʌ/-/ɑ:/, and /ɔ:/-/ɒ/. Some pairs, such as /ɛ/-/æ/, /u:/-/ʊ,/ and /ɑ:/-/ɒ/, were pronounced similarly. Besides, they also can differentiate long and short vowels in the /ɜ:/-/ʌ/, /u:/-/ʊ/, /ʌ/-/ɑ:/, and /ɑ:/-/ɒ/ pairs, but not /i:/-/ɪ/, /ɛ/-/æ/ and /ɔ:/-/ɒ/.

The forth study was conducted by Irawati in 2024, entitled 'Students' Ability in Pronouncing Short and Long Vowel at the Fourth Semester of English Department UIN Syahada Padangsidimpuan.' This study finds out students' pronunciation skills in short and long vowels pronunciation. This study used descriptive quantitative research. The population was the fourth semester students from English Department of State Islamic University Syekh Ali Hasan Ahmad Addary Padangsidimpuan. The instrument of the research was a monolog test. The findings showed that students' skills in pronouncing short and long vowel were categorized into a low level. The students had difficulties in pronouncing short vowel sounds such as /ʊ/ and /æ/, and also long vowel sounds such as /u:/ and /ɔ:/.

The fifth study was conducted by Grima in 2023, entitled 'An Analysis of Students' English Pronunciation Ability at English Department of the Faculty of Arts, Al-Asmarya Islamic University.' This study has two purposes. The first purpose is determining the ability of female students in pronouncing T, linking, and reductions and the second purpose is giving some suggestions and recommendations that might be useful for experts and learners for improving pronunciation. This study is descriptive and analytical

research. The data were from female students' speaking assessment at faculty of Arts in Al Asmarya Islamic University. It was analysed using speech assessment model introduced by Lucas (2012) and Gregory (1990). The scores were computed by utilizing SPSS to obtain the outcomes. The result of the study revealed that the English speaking skill of the students categorized as low in pronunciations T, linking and reductions in the sentences level. The average score was - .64791- 80%. It means the female students need to improve, train, and encourage themselves in pronunciations

The sixth study was conducted by Utin in 2023, entitled 'The Pronunciation of English Long Vowel Sounds by Final Year Students of English of Akwa Ibom State University." This research focuses on examining the English long vowel sounds pronunciation of English students. This study adopted a theoretical framework called Generative Phonology that was introduced by Noam Chomsky and Morris Hale in 1950. The researcher uses a purposive random sampling. The data were collected from a reading test of fifty final year students of English Department in Akwa Ibom State University. The results showed that the easiest sound pronounced by respondents was /ɔː/ and the most difficult one was /ɜː/.