

## **CHAPTER II**

### **LITERATURE REVIEW**

This chapter presents about some theories related to the research. It is about the concept of pronunciation, teaching and learning pronunciation, artificial intelligence in ELT and a review of previous research.

#### **A. Pronunciation**

##### **1. Definition of Pronunciation**

Pronunciation is one of the important aspects of English, especially in oral communication. Every sound, stress, pattern, and intonation may convey meaning. The non-native speakers of English who speak English have to be very careful in pronouncing some utterances or they may create misunderstanding. So, having an intelligible pronunciation is necessary rather than having native-like pronunciation. (Harmer, 2007) defined pronunciation as the way the sounds of a language are made, as the way the sounds and where the word stress and sentence stress are placed, and the way how pitch and intonation are used to indicate our feeling and our meaning.

According to (Hornby, 2008) pronunciation is the way in which a language or a particular word or sound is spoken. Pronunciation is very important to learn, especially in speaking skill. It refers to the way people produce the sounds of words. However, many students find pronunciation to be one of the most difficult aspects in learning English.

In general, pronunciation is the act or result of producing the sounds of speech, including articulation, stress, and intonation, often

concerning some standard of correctness and acceptability.

The study of pronunciation learning is supported by several grand theories in linguistics and cognitive science. One of the most widely recognized is the Speech Learning Model (SLM) proposed by (Flege, 1995). This theory suggests that second language (L2) learners perceive and produce foreign language sounds based on their first language (L1) phonetic system. According to SLM, learners can still acquire new sounds throughout their lives, but their ability to do so is influenced by the extent of their exposure and interaction with the target language.

Additionally, the Communicative Language Teaching (CLT) approach by (Richards & Rodgers, 1986) emphasizes the importance of intelligibility in pronunciation over perfection. This approach suggests that pronunciation should be taught in a way that enhances effective communication rather than merely mimicking native speakers.

## 2. Elements of pronunciation

A broad definition of pronunciation includes both supra-segmental and segmental aspects. “The segmental aspect refers to individual sounds in pronunciation, specifically vowels and consonants, which are essential for distinguishing meaning in spoken language (Yates, 2002). The Suprasegmental Aspect Encompasses elements of speech beyond individual sounds, such as intonation, stress, phrasing, timing, and rhythm, which contribute to the overall tone and flow of speech (Yates, 2002).

### a. Segmental Aspect of Pronunciation

According to (Otlowski, 1998), segmental is individual sounds

of a language. Individual sound is separated into two terms, consonant and vowel. Traditional approaches to pronunciation have often focused on segmental aspects. It is because these traditional approaches relate with the letter in writing term (Yates, 2002). The following is an explanation of the pronunciation aspect.

#### 1) Vowels

Vowels in English can be defined as sounds made by the free flow of air through the vocal tract without resistance or friction. They are characterized by the position of the tongue, lips and oral cavity during speech. Vowels can be a single sound (monophthong) or a combination of two sounds (diphthong). In English there are a number of vowels that play an important role in distinguishing the meaning of words. These vowels can be beaten according to their phonetic properties such as pitch, tongue position, and lip shape. Here are some examples of vowel sounds:

##### a) Short vowels

The standard pronunciation of vowels, usually when there is only one vowel in the word. For example, "a" as in "cat", "e" as in "wet", "i" as in "big", "o" as in "clock", and "u" as in "but".

##### a) Long vowels

For example, "/i:/" as in "week", "/ɑ:/" as in "hard", "/ɔ:/" as in "fork", "/ɜ:/" as in "heard", and "/u:/" as in "boot".

##### b) Diphthongs

A speech sound formed by combining two vowel

sounds to create a new speech sound. Diphthongs are long and complex vowels that start with the sound quality of one vowel and end with the sound quality of another for examples of diphthongs in English include:

- 1) "ai" in the word "pain" (sick)
- 2) "ei" in the word "weight" (weight)
- 3) "oi" in the word "coin" (coins)
- 4) "ou" in "house"

c) A monophthong is a vowel sound that is pronounced with a single tone and mouth position. A monophthong is a vowel sound that can be considered pure. Examples of monophthong sounds in English

/ɪ/: As in "sit"	/ɒ/: As in "pot"
/ʊ/: As in "boot"	/i:/: As in "need", "beat"
/e/: As in "bed"	/ɛ/: As in "went", "bread"
/ə/: As in "letter"	/ɛ:/: As in "care"
/æ/: As in "rat"	/a/: As in "cat", "hand"
/ʌ/: As in "bus"	/ɛ:/: As in "care", "there",

## 2) Consonants

Consonant are key in the human vocal system causing congestion or narrowing of the vocal tract during sound production. Unlike vowels, consonants involve obstructions or obstructions that stop or restrict the flow of air. The main characteristic of consonants

is the presence of obstructions or narrowing of the airflow when spoken. This disability can occur in various ways, e.g. by narrowing of the space between the tongue and palate on /s/ (as in "sit"), complete cessation of airflow by the lips or tongue on /p/ (as in "sit" "pin"), or vibration of the vocal cords with the lips or the tongue for the /z/ sound (as in "zip").

Consonants in English can be classified according to various criteria, e.g. B. after the formation of barriers, the sound and no vibration of the vocal cords.

Here are some categories of consonants in English:

- a) Voiced and voiceless consonants: Voiced consonants have vocal cord vibrations, while voiceless consonants do not have vocal cord vibrations. For example, /p/ is a voiceless consonant (as in "pin"), while /b/ is a voiced consonant (as in "bin").
- b) Explosive consonants: Explosive consonants imply a pause followed by an explosive burst of air. Some examples of consonants that appear in English are /p/ (as in "pin"), /t/ (as in "ten"), and /k/ (as in "cat").
- c) Frictional consonants: Frictional consonants imply resistance which is not followed by explosive bursts of air. For example, /s/ is a frictional consonant (as in "sit") and /f/ is a frictional consonant (as in "fitting").
- d) Nasal consonants: nasal consonants mean the flow of air through the nose. For example, /m/ is nasal (as in "man"), /n/ is nasal (as in

"nut").

- e) Fluid consonants: Liquid consonants include a gentle flow of air without significant resistance. For example, /l/ is a fluid consonant (as in "let") and /r/ is a fluid consonant (as in "run").
- f) Semi-vowel consonants: Semi-vowel consonants have similar properties to vowels, but are still classed as consonants. For example, /j/ is a semi-vowel (as in "ja") and /w/ is a semi-vowel (as in "win"). Consonant pronunciation is very important for understanding and clarity of communication in English. Mispronouncing a consonant can change the overall meaning of a word or make it difficult for the listener to understand it. Therefore, it is important for English learners to practice and have a good understanding of pronouncing consonants, including how barriers form, how the tongue moves, and air escaping through the vocal tract.
- g) voiceless: Airflow from the lungs continues down the trachea or trachea and passes through the opening/cavity in the vocal cords which is called the vocal cleft. When the vocal cords are separated/absent, airflow is unobstructed through the auditory canal and freely enters the supraglottic cavity (the part of the vocal organ above the auditory canal).

Called – voiced: For example, p, t, k, and s in English, the words pit, tip, kit, sip, and kiss are silent. When the vocal cords come together, the airflow direction is fixed and vibrates. To illustrate,

these sounds are called b, d, g, and z in the words Angry, God, Dog, Zebra, and Buzz.

b. Suprasegmental aspects

Suprasegmental aspects of pronunciation are phonetic features of speech that affect units larger than individual sounds, such as syllables, words, or phrases. Suprasegmentally is also called prosody. Suprasegmental features include: intonation, stress, phrasing, and rhythm. Suprasegmentally plays an important role in conveying meaning and emotion in spoken language. Native speakers can automatically understand the differences in grammar, emotion, and continuity marked by suprasegmental. Below is a more complete explanation of the suprasegmental aspect.

1) Intonation

Intonation describes how the voice rises and falls in speech. The three main patterns of intonation in English are: falling intonation, rising intonation and fall-rise intonation.

a) Falling Intonation (↘)

Falling intonation, where the pitch of the voice drops at the end of a sentence, is the most common intonation pattern in English. It typically occurs in statements, commands, WH-questions (also known as information questions), confirmatory question tags, and exclamations.

Examples of falling intonation:

### Statements

- Nice to meet ↘you.
- I'll be back in a ↘minute.
- She doesn't live here ↘anymore.
- Dad wants to change his ↘car.
- Here is the weather ↘forecast.

### Commands

- Write your name ↘here.
- Show me what you've ↘written.
- Leave it on the ↘desk.

### Wh- questions (requesting information.)

(questions beginning with 'who', 'what', 'why', 'where', 'when', 'which', and 'how')

- What country do you come ↘ from? Where do you ↘ work?
- Which of them do you ↘ prefer? When does the shop ↘ open?
- How many books have you ↘ bought?

Questions Tags that are statements requesting confirmation rather than questions. Not all tag questions are really questions. Some of them merely ask for confirmation or invite agreement, in which case we use a falling tone at the end.



For example:

- He thinks he's so talented, doesn't ↘ he?
- She's always so helpful, isn't ↘ she?

Exclamations also commonly employ a falling intonation to convey strong feelings or reactions:

- What a wonderful surprise ↘ this is
- ! That's exactly what I ↘ wanted!

On the other hand, rising intonation (↗), where the pitch of the voice increases towards the end of a sentence, is typically used to prompt a response or indicate a genuine question. This intonation pattern is commonly associated with yes/no questions and tag questions that function as real inquiries.

For example:

- Did you complete your assignment ↗ already?
- Can I use your ↗ notebook?
- May I join your ↗ group?

The use of intonation plays a crucial role in determining the meaning and intention behind spoken language, particularly in differentiating between statements seeking confirmation and genuine questions.

Questions tags that show uncertainty and require an answer (real questions).

- We've met already, ↗ haven't we? You like fish, ↗ don't you?
- You're a new student ↗ aren't you? The view is beautiful, ↗ isn't it?

#### b) Rise-Fall Intonation (↗↘)

(The intonation rises and then falls.)

Rise-fall intonation is used in choices, lists, unfinished thoughts, and conditional sentences.

##### 1. Choices (alternative questions):

- Are you having ↗soup or ↘salad?
- Is John leaving on ↗Thursday or ↘Friday?
- Does he speak ↗German or ↘French?
- Is your name ↗Ala or ↘Ali?

##### 2. Lists (rising, rising, rising, falling):

Intonation rises on each item and falls on the last to indicate completion.

- We've got ↗apples, ↗pears, ↗bananas, and ↘oranges.
- The sweater comes in ↗blue, ↗white, ↗pink, and ↘black.
- I like ↗football, ↗tennis, ↗basketball, and ↘volleyball.

##### 3. Unfinished thoughts (partial statements):

The rise-fall pattern indicates hesitation or reservation.

- Do you like my new handbag? → Well, the ↗leather is  
↘nice... (but I don't like it).
- What was the meal like? → Hmm, the ↗fish was  
↘good... (but the rest wasn't great).
- So you both live in Los Angeles? → Well, ↗Alex  
↘does... (but I don't).

#### 4. Conditional sentences:

The first clause typically has a rising tone; the second has falling tone.

- If he ↗calls, ask him to leave a ↘message.
- Unless he ↗insists, I'm not going to ↘go.
- If you have any ↗problems, just ↘contact us.

#### c) Fall-Rise Intonation (↘↗)

(The intonation falls and rises within a word or phrase.) Fall-rise intonation is commonly used to indicate uncertainty, hesitation, or politeness.

##### 1. Hesitation or reluctance:

Used when the speaker is unsure or unwilling to respond directly.

- So you'd be willing to confirm that? → Well... I  
↘sup↗pose so...
- You didn't see him on Monday? → I don't quite

ˈreːmber...

## 2. Polite suggestions or doubt:

Shows tentativeness or politeness in offering an idea.

- Perhaps we could ˈvɪzɪt the place?
- Should we ˈkɒpi the list?

## 2) Stress

Stress is one of the elements needed to be learned. By looking at the English dictionary, the English learners can identify the stress. Stress occurs when a syllable is given an additional force of energy (Roach, 2012). The stress is shown in every word and phrases by putting the symbol of an acute accent (‘) in the syllable sound. The syllable sound appears when the word is pronounced. To identify the syllable sound is simple by identifying the vowel sound. One vowel sound represents one syllable sound. The vowel sound is monophthong, diphthongal, and diphthongal sound. Thus, before the learners learn how to stress the English words, they must know what the syllable is.

A syllable is a unit of speech that consists of one or more phonemes and is made up of a vowel or vowel-like sound that may be preceded or followed by consonants (Kreidler, 2004). One syllable implements only one vowel sound. To identify it, the syllable is made of phonemes. ‘book’

/bɒk/ has three phonemes; consonant /b/, vowel /ɒ/, and consonant /k/ as it is a one-syllable-sound word. There are three types of stress.

- a) Primary Stress: The syllable with the Primary Stress is the loudest and longest syllable in a word. It has a very noticeable change in pitch on the vowel. In one-syllable words, that one syllable gets the primary stress. Examples:

“blue” =

/BLUUU/

“ten” =

/TEHN/

The only exceptions to this rule are a handful of function words like “the” which are usually unstressed or reduced.

- B. Secondary Stress: These are the syllables that are stressed, but not *as* much the primary stress, so they are loud and long with a change in pitch, but not as loud or as long as the syllables that get the primary stress. The change in pitch on the vowel is not quite as noticeable.

Example:

“disappear” = /DIH-suh PEEER/ (syllable with secondary stress is underlined)

- C. Level 3 is the completely Unstressed Syllable: These are the syllables that have no stress at all so we rush through them and shorten them so much that the vowel in the syllable is almost

completely gone. The change in pitch on the vowel is barely noticeable, but it is there.

Example: “prepare” = /pruh-PAIR/ (unstressed syllable is underlined)

### 3) Phrasing

Phrasing in pronunciation refers to how speakers group words together in meaningful units or "phrases" to convey ideas smoothly and naturally. Phrasing is closely tied to intonation, stress, and pausing in speech and is essential for effective communication, as it affects both the clarity and the flow of spoken language.

#### Key Aspects of Phrasing

1. **Chunking:** Breaking speech into manageable parts, or "chunks," helps listeners process and understand the information more easily. This usually involves grouping words that naturally belong together, like noun phrases, verb phrases, or prepositional phrases.

Example: "I went to the store // to buy some groceries."

2. **Pausing:** Pauses are used between phrases to signal the end of one idea and the beginning of another. Effective pausing can prevent speech from sounding rushed or muddled.

Example: "After a long day at work, // she decided to relax."

3. **Intonation:** Intonation refers to the rise and fall of pitch

within a phrase. It guides the listener through the speaker's intent (whether it's a statement, question, or exclamation) and highlights important words. Example: In a question, the pitch often rises towards the end: "Are you coming with us?"

4. Stress within Phrases: Certain words within a phrase receive more emphasis to convey the main point or meaning. In English, content words (nouns, verbs, adjectives) are often stressed, while function words (like articles and prepositions) are typically unstressed. Example: "She wants to go to the party."

#### 4) Rhythm

In phonetics, rhythm is the speed and cadence of how a sentence is spoken, and is marked by the quantity, timing, and stress of syllables. It's a sense of movement in speech that's concerned with syllables and larger parts of speech.

Rhythm is important for making speech sound flow well and for helping people understand what's being said. Examiners consider rhythm, stress, and intonation when evaluating pronunciation. These are important markers of fluency and can change the meaning of a word or sentence. The two most common types of rhythm in language are stress-timed rhythm and syllable-timed rhythm. English typically uses a stress-timed rhythm, which relies on the regular, repetitive occurrence

of stressed and unstressed syllables.

Example: Peter Piper picked a peck of pickled peppers. Below, the stressed syllables are in bold:

**Peter Piper picked a peck of pickled pepper**

#### 1. Factors Influencing Students English Pronunciation

The learning of pronunciation is influenced by a number of important aspects. (Brown, 2005) cites (Kenworthy, 1987) as summarizing the key learner factors that affect pronunciation mastery. These include motivation and care for proper pronunciation, age, exposure, intrinsic phonetic ability, identity and language ego, and native language. learning.

- a. Native language: It is clear that the main factor influencing students' pronunciation is the crucial role of the mother tongue. All languages in the world contain a wide variety of phonological elements. These variations can cause difficulties for learners who want to pronounce words well in their target language. Therefore, the recognition and effort of the learners who have been done to overcome these difficulties can help improve English pronunciation.
- b. Age: Generally, up until adolescence, children can maintain perfect or native pronunciation in a foreign or second language, provided they are constantly exposed



to authentic contexts. Once adolescence, also known as the critical period, is reached, normal language acquisition becomes impossible. However, adults can also reach a local accent, as children do. A study conducted by Singleton (2005) has also reached the conclusion that it is not clear if the hypothesis of the critical period can probably be considered as a logical hypothesis to obtain the perfect native pronunciation of many disputes associated with this critical period.

- c. Exposure: The social environment is very important as it gives students a great opportunity to become familiar with the target language. The language learning theories of Postovsky (1974), Asher (1977), and Krashen (1982), among others, posited that learners acquire language through input, and they must receive multiple comprehensible inputs, defined as “everything about the L2/target language that learners encounter” (Saville- Trolke, 2006, p. 74) before they can speak. If this is true, then exposure to the target language is a determining factor in students' success: students must have ample opportunities to receive appropriate feedback that they can understand. They need to expose themselves to target language input through opportunities to experience authentic oral

discourse samples by native speakers of the target language both inside and outside the classroom, such as in language laboratories and learning centers.

- d. Natural voice ability: Natural voice ability is sometimes called voice coding ability (Brown, 1992).

The general point of view is that some language learners can make more accurate sounds and imitate better sounds than other sounds. Students who are born and live in an English-speaking country are often more likely to acquire and experience native-like pronunciation than students who are not born and live in an English-speaking country. However, because learners are born with different phonetic abilities, biology, and physiology, not all learners who are born and live in a native-speaking environment will consistently achieve native-speaker pronunciation. As a result, their success, which sounds as an indigenous person, is determined mainly by their daily impact on target languages.

- e. Identity and the linguistic ego: The attitude of speakers towards the target language and the amount to which the linguistic ego identifies with these speakers is also another crucial factor impacting pronunciation learning. A speaker's ability to establish their identity

and membership to specific socio-cultural contexts through language is crucial, as demonstrated by their positive interactions with other speakers of the target language. More significantly, linguistic self is essential for assisting pupils in developing native-like pronunciation and being self-reliant and self-assured when speaking English. Learners who have an open mind and a positive attitude toward the target language and target language speakers are more likely to pick up and improve their pronunciation more quickly, according to a substantial body of research on the relationship between pronunciation, identity, and linguistic self. More effectively (Dörnyei, 2003; Brown, 1992; Ahmed, 2017)

f. Motivation and interest in good pronunciation: Of the six factors, motivation and interest in proper pronunciation are arguably the most crucial. If a learner is intrinsically motivated, they may be able to get native-like pronunciation; if they are not, they might not be motivated to perform well. According to certain research, motivation is crucial for learning proper pronunciation (Luchini, 2006; Meléndez, 2006; Tanner, 2012; Ute & Christiane, 2000).

g. Confidence: Confident students are more likely to

practice speaking and take risks with their pronunciation.

- h. Listening Skills: Good listening skills enable students to pick up on nuances in pronunciation and intonation.

## 2. Teaching and Learning Pronunciation.

(Harmer, 2007) defines pronunciation as how sounds are produced, how stress is applied, and how intonation and pitch are used to convey meaning. The definition of additional pronunciation from Oxford's vocabulary is a way to speak language or specific words or sounds. When referring to the "correct pronunciation", it can apply to both in a given dialect. Words may be pronounced differently by different people or groups of people depending on many factors such as where they grew up, where they currently live, whether they have a language or speech disorder, the group's ethnicity, social class or education. Teaching pronunciation is teaching about aspects that influence the meanings of sentences through segmental phonemes, this involves not only the accurate production of individual sounds but also the mastery of suprasegmentally features such as stress, intonation, and rhythm, which are crucial for conveying the intended meaning and emotion in spoken language. Therefore, the method of teaching pronunciation must also be correct because if it is wrong, many mistakes will occur in learning English.

In teaching pronunciation there are many approaches or

methods used such as tongue twister, drilling, direct method and audio lingual method in this study, an application was developed to be used as a medium for learning English pronunciation independently where the application emphasizes the parameters of emphasizing sound pronunciation and comparing it with the pronunciation in the database so that it can produce a value for the accuracy level of the pronunciation given.

Students face several challenges in learning pronunciation, including issues related to the teacher's instruction, accent, intonation, word stress, and motivation, Elliot (1995) concluded that students who focused more on their pronunciation were able to pronounce English as a second language more accurately. According to him, pupils who were motivated to learn and had a positive mindset regarding the target language and its speakers achieved better than those who had negative sentiments.

Gardner and Lambert (1972) stated that students with integrative motivation are more likely to make greater efforts to improve their second language communication skill.

Attitudes toward language learning play a significant role in students' success. As cited in Despagne (2003), Dornyei explained that attitudes are created, and can be changed, through family, teachers, peers, and school. This means that learners' environments strongly influence their motivation and beliefs

about learning. Attitudes consist of three components: cognitive, affective, and behavioral (or conative). The affective component refers to emotional reactions. For instance, “I’m afraid of spiders” reflects a person's feelings. The behavioral component reflects how these attitudes influence actions, such as “I will avoid spiders and scream if I see one.”

In the context of English learning, many students struggle with pronunciation due to negative attitudes. Their lack of interest, low motivation, and belief that learning English is not important become major obstacles. These attitudes reduce their concentration and willingness to practice, especially in improving their pronunciation. Moreover, many of them have limited perspectives regarding how English could benefit their future.

Teacher-related pronunciation problems are a significant part of classroom language practice. According to Morley (1991), teachers help students learn pronunciation rather than merely instructing them. They act more like coaches—providing guidance, correction, and support throughout the learning process. Similarly, Harmer (2007) argues that when it comes to teaching pronunciation, most teachers focus more on teaching vocabulary and grammar than on teaching pronunciation. Harmer (2007) stated, "Almost all English language instructors encourage their students to learn vocabulary and grammar,

engage in productive skill activities, practice functional conversations, and work toward becoming proficient readers and listeners”. However, some of these same professors merely briefly address pronunciation and make no effort to teach it clearly. This indicates that teachers make little attempt in teaching pronunciation because their accents are not like native speakers. Additionally, pupils' pronunciation of English sounds may be affected by the pronunciation of their English Profession. Regarding pronunciation teaching, Harmer (2007) recommends a number of approaches such as separating similar and confusing phrases utilizing word pairs, introducing stress, and practicing intonation. These concepts suggest that teaching pronunciation is essential for English language instructors and that teaching other subjects has been ignored for the purpose of teaching pronunciation.

In their study on the acquisition of English, (Marinova-Todd et al. 2000) came to the conclusion that, with enough motivation, second language learners can become extremely proficient—even native-like speakers. According to Bernaus et al. (2004), the need and desire for native-like pronunciation might be influenced by one's personal or professional objectives for learning English. Additionally, Moyer (2007) discovered that positive orientation and experience are key components in the development of native-like pronunciation. Both current learning

and previously learned abilities and methods can be impacted by student motivation (Schunk et al., 2008). Motivation is the length that pushes the person to achieve a goal; if the learners do not have motivation the learning process is harder. It cannot be denied that motivation is an important factor that determines the people's participation in every activity especially in language learning. (Brophy, 2010) described student motivation as the extent to which learners invest time, energy, and attention toward various tasks, goals, or activities.

Word stress is a crucial aspect of English pronunciation, and correct intonation plays a vital role in comprehending spoken English, because its importance in comprehending spoken English, word stress is an essential component of English pronunciation. English native speakers naturally employ word stress, which may cause non-native speakers to find talks challenging. The "shape" of a word or phrase is largely determined by its stress, and the wrong stress can confuse conversations.

When syllables are not stressed by any norms, communication issues arise for both native speakers and English language learners. There are five tone marks and six tones in Vietnamese, which alter the meaning of words. In English, intonation has a significant significance, but Vietnamese learners typically do not pay attention to learning English intonation due



to a lack of awareness regarding the importance of correct intonation.

### 3. Technology in Language teaching and learning

In recent years, technology has become a critical component in language teaching, transforming how pronunciation is taught and learned. With advancements in digital tools and resources, learners now have greater access to pronunciation practice, individualized feedback, and authentic language models. These innovations make it possible for language learners to overcome some of the traditional challenges associated with pronunciation learning, such as limited access to native speakers and the difficulty of accurately perceiving and producing unfamiliar sounds.

#### a. Access to Native and Authentic Models.

One of the most significant advantages of technology in pronunciation learning is the availability of high-quality native or near-native models. Unlike earlier times, when learners had to rely on teachers or textbook recordings, modern technology provides various resources to listen to accurate pronunciation. Mobile apps like Duolingo, Babbel, and Rosetta stone offer native recordings for vocabulary and sentence structures, helping learners improve their listening and speaking skills. Online platforms like YouTube and podcast services also enable learners to engage with authentic language content, allowing them to mimic the pronunciation, intonation, and rhythm used by native speakers.

Additionally, technology facilitates exposure to different dialects and accents. Learners can listen to British, American, Australian, or other English accents and become familiar with the global diversity of the language. This exposure would have been challenging to achieve in traditional classroom settings where learners typically interact with a single teacher or accent.

b. Pronunciation Feedback through Speech Recognition and AI.

Speech recognition technology has also made it possible to provide learners with real-time feedback on their pronunciation. Advanced AI- driven apps like ELSA Speak and Google Assistant use sophisticated algorithms to analyze a learner's speech and compare it to native speaker models. These tools can pinpoint specific sounds, words, or intonation patterns that the learner needs to improve, offering corrective feedback that helps learners refine their pronunciation over time.

For example, speech recognition systems can highlight problematic phonemes, such as the difficulty in distinguishing between the "r" and "l" sounds for some learners, and provide targeted practice drills. AI algorithms can also track a learner's progress over time, creating personalized learning plans to address the unique pronunciation challenges faced by individual learners.

This level of customization enhances the efficiency of pronunciation learning, allowing students to focus on their

specific weaknesses and make noticeable improvements. Self-Paced Learning and Autonomous Practice.

The integration of technology into language learning has also empowered learners to practice pronunciation at their own pace, independently of classroom constraints. Interactive language learning platforms offer features like repetition exercises, voice recordings, and immediate playback options, enabling learners to practice speaking and self-correct their pronunciation. Tools like Forvo or Mango Languages provide pronunciation guides where learners can hear a native speaker's pronunciation, record their version, and compare the two, allowing for self-assessment and improvement.

Moreover, gamified apps make pronunciation practice more engaging. By turning learning into a game with points, levels, and challenges, platforms motivate learners to dedicate more time to pronunciation drills, leading to more consistent practice. For learners who may feel self-conscious about their pronunciation in a group setting, technology provides a safe and private environment where they can practice without fear of embarrassment.

c. Virtual Reality (VR) and Augmented Reality (AR)

Emerging technologies like virtual reality (VR) and augmented reality (AR) are also being explored for pronunciation learning. With VR, learners can immerse

themselves in virtual environments where they can interact with native speakers or virtual characters, simulating real-life conversations.

For instance, VR applications create scenarios such as ordering food in a restaurant or asking for directions, where learners must pronounce words clearly to continue the interaction. This immersive experience can boost learners' confidence and fluency, as they practice speaking in context and receive immediate feedback from the system or virtual interlocutors.

Similarly, AR can be used to augment the physical world with interactive pronunciation guides. For example, learners can use AR apps to point their smartphones at an object and see how it is pronounced in the target language, with audio and visual cues that aid memory retention. These immersive experiences help bridge the gap between traditional learning and real-world application.

#### d. Overcoming Pronunciation Challenges with Technology

Despite the clear advantages, technology in pronunciation teaching is not without its limitations. Automated speech recognition systems may sometimes misinterpret non-native accents or fail to provide feedback on subtler elements like stress and intonation. Furthermore, while self-paced learning is beneficial, some learners may struggle without the guidance of

an expert teacher who can provide more nuanced feedback and instruction on how to shape sounds or correct errors.

Nevertheless, when used alongside traditional instruction, technology can significantly enhance the teaching and learning of pronunciation. It provides learners with the resources, flexibility, and immediate feedback they need to develop their pronunciation skills effectively.

## **B. Artificial Intelligence in English Language Teaching**

AI, Artificial Intelligence, is the term used to describe the development of computer systems that are capable of performing tasks that would typically require human intelligence, such as understanding natural language, identifying patterns, coming to methodical conclusions, solving issues, etc. Building computers with human intellect that can learn from experience and process, analyze, and use data to improve their performance over time is the foundation of artificial intelligence (AI) technology (Laupichler et al., 2022). These days, artificial intelligence (AI) is employed in many different fields, such as language translation, picture processing, speech recognition, and self-driving cars.

The development of systems that can accurately and efficiently carry out human-like tasks is the ultimate goal of AI research (Bartneck, 2021) AI has the potential to enhance many facets of human existence today, including healthcare, education, entertainment, and transportation. To be more specific, AI-powered platforms or applications are typically utilized in education to give students real-time feedback and personalized learning experiences

(Luckin et al., 2022). They are learning more efficiently and at their own speed thanks to this. AI that can evaluate student data can pinpoint areas in which a student needs more assistance and then offer resources and direction in those areas.

The use of AI technology to teach English is known as "AI in English Language Teaching" (ELT). Personalized learning programs, chatbots, language learning games, and virtual tutors are all part of this application. Until now, the history of AI in ELT can be traced back to the early years of computer-assisted language learning (CALL) in the 1970s and 1980s, when computer technology was initially employed to support language instruction (Rukiati et al., AI on Learning English: Application, Benefit, and Threat 33). The goal at the time was to create a grammar checker (Schank, 1987).

Initially, language learning software primarily focused on drilling vocabulary and grammar. However, with the advancement of Natural Language Processing (NLP) and Artificial Intelligence (AI) in the 1990s and 2000s, language learning technologies began to incorporate more interactive and adaptive features. These innovations enabled systems to simulate human-like interactions and provide personalized learning experiences. For instance, intelligent tutoring systems began offering customized feedback and adapting content based on individual learners' performance (Woolf, 2009; Heffernan & Heffernan, 2014). Research has shown that AI-powered tools can improve student engagement and learning outcomes by responding dynamically to learners' needs (VanLehn, 2011).

Today, AI is increasingly utilized in English language teaching (ELT)

to make language learning more engaging, efficient, and accessible. AI-driven language learning applications and platforms such as ChatGPT, Coursera, Duolingo, ELSA, Grammarly, among others, have gained popularity among language learners. AI is also being used to assist language teachers in their instruction. As AI technology continues to evolve, it is anticipated that AI will play an even more crucial role in ELT by offering more personalized experiences and providing effective language learning support for students worldwide (Yoon, 2019).

Some AI-based educational applications and platforms also use NLP to create virtual tutors that can engage in conversations with students and help them improve their language skills in a more conversational environment. An example of AI in ELT is a virtual English tutor that uses NLP and machine learning algorithms to assess a student's speaking and writing abilities and offer personalized feedback (Lu, 2018). This virtual tutor can also generate practice exercises and activities tailored to the student's needs.

### **C. Artificial Intelligence in Pronunciation**

AI has significantly transformed autonomous English learning by providing personalized and adaptive learning experiences that enhance speaking and listening skills, improve accessibility, and offer valuable feedback. For instance, AI-powered applications in ELT equipped with speech recognition technology can assess students' pronunciation, deliver real-time feedback, and support learners in enhancing their speaking abilities. Such platforms can analyze students' performance data, recommend suitable lessons, and offer instant corrections related to pronunciation, grammar, and

vocabulary. Additionally, these tools provide interactive activities that allow students to engage with learning materials in innovative ways. By empowering learners to take charge of their learning journey, AI aims to make language learning more efficient, engaging, and accessible to students worldwide (Han, 2019).

#### **D. Review of Previous Research**

In this research, there are some reviews of related literature finding from the previous researcher, they are:

The first previous research finding come from (Kholis, 2021) in His research entitled *“Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills”* Their research focused on enhancing English pronunciation skills among higher education learners, specifically English Department students, through the use of the ELSA Speak App. They emphasized that in speaking, proper pronunciation is essential, as it involves mastering phonetics and phonology to accurately articulate words and produce sounds. Clear pronunciation is crucial for conveying messages effectively and ensuring they are understood by the listener.

This research conducted to know the effect of using The ELSA Speak app in supplementing the students’ pronunciation skills. The researcher used this app in learning to pronounce for the reason that it was the one automatic speech recognition used by most and contribution in enhancing pronunciation skill. The similarity of research between previous researcher and current research is the first in topic raise namely pronunciation, the second refers to the object of research students. The difference between



the two studies is that previous research concentrated on the implementation of the ELSA Speak Application to enhance students' pronunciation skills, while the current study focuses on facilitating pronunciation in the context of storytelling.

The second previous study conducted by Lestari, R. et al. (2024) with the title “*The Impact of Using the English Pronunciation Digital Application on Students' Pronunciation Proficiency at MAN Labuhanbatu, North Sumatra*” showed that the use of digital applications has a significant positive effect on improving students' pronunciation skills. The study found that before the intervention, students had difficulty in pronunciation accuracy, but after using the app, their ability improved significantly. This confirms that the integration of digital tools in language teaching is very effective in supporting the development of pronunciation skills.

These findings are relevant to the research entitled “The Effectiveness of ELSA Speak Application in Facilitating Students' English Pronunciation in Story Telling”. The difference is that my research focuses on the context of narrative text learning through storytelling practice. Besides aiming to improve students' pronunciation, this study also explores how an AI-based pronunciation application can help students in developing their storytelling skills, while providing convenience for teachers in teaching. The Researcher complements previous findings by adding the dimension of storytelling practice as part of strengthening students' communication skills.

The previous research were conducted by (Adawiah & Muliati, 2024) entitled “The Effect of ELSA Speak on Pronunciation of English Speaking at

Vocational High School". This study aims to determine the effectiveness of the ELSA Speak application in improving students' pronunciation skills at the vocational high school level. Using a one-group pretest-posttest pre-experimental design, the researcher involved 13 students from the XI Fashion class of SMK Sritanjung as samples. The results showed that there was a significant increase in students' average score from 43.31 in the pretest to 85.23 in the posttest, with a significance value of 0.000 ( $<0.05$ ). In addition, correlation analysis showed a moderately strong relationship ( $r = 0.651$ ) between the pretest and posttest results, indicating that the improvement was closely related to the intervention using ELSA Speak. This study indicates that the ELSA Speak app is effective in helping students improve English pronunciation independently through features such as phoneme training, intonation, syllable stress, and real-time feedback. This study strengthens the argument that AI-based technology can be an effective learning tool to improve students' speaking skills, particularly in terms of pronunciation.

However, there are some research gaps that need to be bridged. This study focused on improving pronunciation technically without linking it to a broader context of language use, such as storytelling. In addition, the approaches in previous studies have not explored how these pronunciation skills are applied in real communication-based activities that can increase students' confidence, such as storytelling. Research has also not utilized artificial intelligence (AI)-based technologies that can provide personalized and adaptive feedback to support pronunciation learning.

Several previous studies have investigated the effectiveness of ELSA Speak in improving students' pronunciation skills. Kholis (2021), in his study

entitled “ELSA Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills”, explored how the app could support English department students in higher education. His findings confirmed that pronunciation, as a crucial component of speaking, can be significantly enhanced through the use of AI-powered tools like ELSA Speak. Similarly, Lestari et al. (2024) found that the integration of digital pronunciation applications significantly improved students' pronunciation accuracy at the senior high school level. Another study by Adawiah and Muliati (2024), conducted at a vocational high school, reported a significant improvement in students' pronunciation after using ELSA Speak, as evidenced by the increase in their post-test scores and a strong correlation between pre-test and post-test results.

Although these studies have proven the effectiveness of ELSA Speak in enhancing pronunciation, they primarily focused on the technical aspects of pronunciation in isolation. There is a noticeable gap in the integration of pronunciation learning within a meaningful communicative context, such as storytelling. Furthermore, none of the studies were conducted at the junior high school (SMP) level, and none examined how AI-based feedback from ELSA Speak supports students' pronunciation performance specifically in narrative text learning. This study addresses those gaps by evaluating the effectiveness of ELSA Speak in facilitating students' English pronunciation within the context of storytelling. By doing so, this research not only aims to improve students' pronunciation but also to enhance their confidence and fluency through communicative storytelling practices, while providing

practical value for teachers in delivering pronunciation instruction more effectively