CHAPTER 2

LITERATURE REVIEW

A. Pronunciation

1. Definition of Pronunciation

Kenworthy (1987) views pronunciation as a bridge connecting the abstract language system with the actual spoken form in oral communication. According to Ur (1996), pronunciation is a speaking technique that does not require a person to completely imitate the accent of a native speaker, but rather focuses on the speaker's ability to pronounce words correctly so that the message conveyed can be easily understood by listeners who have good language skills. Kelly (2000) states that pronunciation is the ability to produce and differentiate the sounds of a language, including vowels, consonants, stress, and intonation, in order to produce intelligible speech. Pronunciation is not only about producing sounds correctly, but also about the ability to convey meaning clearly.

Furthermore, pronunciation involves various instrumental components in sound production. From a phonological perspective, these components can be analyzed into two main complementary categories. The first category is segmental features, which focuses on individual sounds such as vowels and consonants. The second category is suprasegmental features, which includes musical aspects of speech such as rhythm, stress, and intonation that extend across these segments.

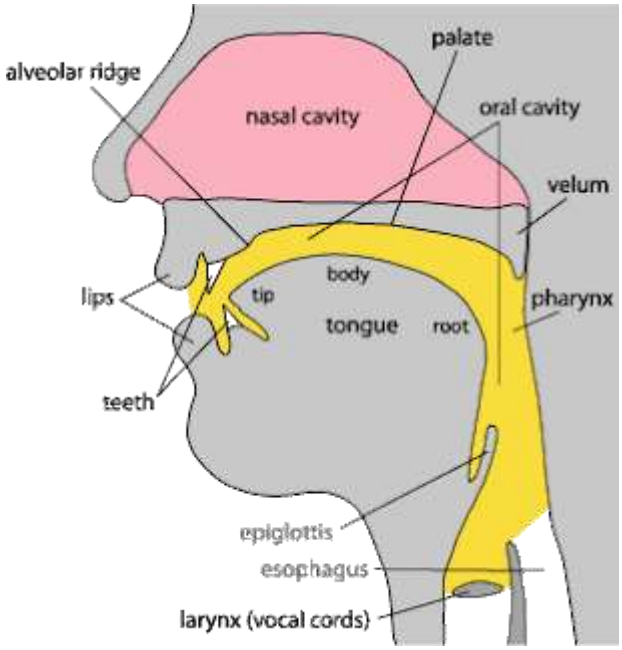
2. Features of English Pronunciation

a. Segmental Features

According to Kenworthy (1987), segmental features are individual aspects of a language's sound system consisting of vowels and consonants as basic sound units or phonemes. Segmental features are separate sounds that can be analyzed individually in word and sentence structures. The focus in segmental learning is on

the identification and production of phonemes and the differences in meaning that arise from differences in a single sound. Vowels are sounds produced without significant obstruction in the vocal tract, while consonants are produced with some obstruction by speech organs (e.g., lips, tongue, teeth, palate). In contrast, consonants are produced by obstructing or narrowing the flow of air with speech organs.

Figure 1. Speech Organs



The quality of vowel sound is determined by the shape of oral cavity, which is controlled by the position of the tongue (high-low, front-back) and the shape of the lips (rounded or not). Here are some examples of vowels w in English:

/ɪː/	bee	/bɪː/
/ɪ/	bit	/bɪt/
/ʊ/	put	/pʊt/
/uː/	food	/fuːd/
/e/	bed	/bed/
/ə/	ago	/əgəuː/

/ɪə/	h<u>ea</u>r	/hɪə/
/eɪ/	s<u>ay</u>	/seɪ/
/ʊə/	s<u>ure</u>	/ʃʊə/
/ɔɪ/	b<u>oy</u>	/bɔɪ/
/əʊ/	l<u>ow</u>	/ləʊ/
/eə/	h<u>air</u>	/heə/
/aɪ/	f<u>i</u>ne	/faɪn/
/aʊ/	n<u>ow</u>	/naʊ/
/ɒ/	h<u>o</u>t	/hɒt/
/ɜː/	w<u>o</u>rk	/wɜːk/
/ɔː/	w<u>a</u>lk	/wɔːk/
/æ/	b<u>a</u>d	/bæd/
/ʌ/	b<u>u</u>t	/bʌt/
/ɑː/	c<u>a</u>r	/cɑː/

Meanwhile, consonant sounds are described based on three main criteria: place of articulation (where the obstruction occurs), manner of articulation (how the air is obstructed), and voicing (whether the vocal cords vibrate or not). Here are some examples of consonants in English:

/p/	<u>p</u>en	/pɛn/
/b/	<u>b</u>all	/bɔːl/
/t/	<u>t</u>en	/tɛn/
/d/	<u>d</u>ay	/deɪ/
/k/	<u>k</u>ey	/kiː/
/g/	<u>g</u>ot	/gɒt/
/f/	<u>f</u>ood	/fuːd/
/v/	<u>v</u>ain	/veɪn/

/θ/	thin	/θɪn/
/ð/	they	/ðeɪ/
/s/	small	/smɔːl/
/z/	zoo	/zuː/
/ʃ/	shell	/ʃɛl/
/ʒ/	vision	/ˈvɪʒᵊn/
/h/	house	/haʊs/
/m/	moon	/muːn/
/n/	no	/nəʊ/
/ŋ/	string	/strɪŋ/
/tʃ/	chair	/ʧeə/
/dʒ/	just	/ʤʌst/
/l/	look	/lʊk/
/r/	real	/rɪəl/
/j/	yes	/jɛs/
/w/	win	/wɪn

b. Suprasegmental Features

Suprasegmental features in English pronunciation include aspects that affect more than one sound segment, such as stress, intonation, and rhythm, which are often referred to as the musical aspects of pronunciation. Celce-Murcia et al. (2010) emphasize that the musical aspects of pronunciation are very important, with some researchers even suggesting that these aspects are more important than the pronunciation of individual sounds. In the classroom context, teachers need to help students practice both individual sounds and overall language patterns.

The following are details of each of the main features and their significance:

1. Word Stress and Sentence Stress

Word stress is the emphasis on one syllable in a word (Celce-Murcia et al., 2010). While, sentence stress is the emphasis on key words in a sentence. Kenworthy (1987) claims that errors in stress placement can make words hard to recognize. So, word stress is often seen as a main priority in pronunciation teaching; Ur, 1996).

2. Rhythm

   Rhythm is the pattern of stressed and unstressed syllables in speech. English is a stress-timed language, while Indonesian is more syllable-timed, so learning the stress-timed pattern helps learners sound more natural in English (Celce-Murcia et al., 2010; Kelly, 2000; Kenworthy, 1987).

3. Intonation

   Intonation is the 'melody' or 'music' in speech produced by the rise and fall of pitch when speaking (Kelly, 2000). Basic intonation patterns in English include falling intonation, rising intonation, and combinations of the two (fall-rise or rise-fall).

   Intonation conveys layers of meaning that are not conveyed by the words themselves. The importance of intonation lies in the following functions:

   a. Grammatical Function: Distinguishing sentence types. For example, a falling pattern usually indicates a statement (I live in Jakarta.), while a rising pattern often indicates a yes/no question (Do you live in Jakarta?).
   b. Attitudinal Function: Conveys the speaker's attitude and emotions, such as enthusiasm, doubt, sarcasm, or boredom. According to Kenworthy (1987), intonation indicates the speaker's level of

emotional involvement in the conversation.

c. Discourse Function: Organizes the flow of information in a conversation. For example, a fall-rise pattern is often used to indicate unfinished or contrasting information, while a decisive downward pattern indicates the end of an idea.

d. Accentual Function: Places special emphasis on new information or information that the speaker considers important.

Errors in intonation can lead to serious misunderstandings at the pragmatic level. Even if the listener understands every word spoken, they may misinterpret the speaker's intention, emotion, or attitude. Therefore, mastery of intonation is vital for achieving communicative competence.

3. Learning pronunciation

The process of learning pronunciation is influenced by various complex and interrelated factors. Kenworthy (1987) identified several key factors including age, motivation, target language attitude, and exposure opportunities. Contemporary research has expanded this understanding, with Jihad et al. (2024) identifying mother tongue interference, phonological awareness, practice habits, and social interaction patterns as significant contributors. Most notably, Arsifa et al. (2024) found that language identity and ego account for approximately 80% of pronunciation proficiency variation among Indonesian junior high school students.

Age is not the only factor that shapes learners' pronunciation. Although Kenworthy (1987) notes that children generally show a stronger ability to adopt second language pronunciation, but age alone does not determine the outcome. Individual effort and goal setting can

contribute to pronunciation accuracy regardless of age, especially through active social interaction with native speakers or proficient users (Jihad et al. 2024). More important than age is the combination with motivation, exposure opportunities, and target language attitudes.

Motivation emerges as the dominant factor in pronunciation development. Ikhsan (2017) found that students developed pronunciation skills when motivated by authentic English media, especially films and songs. Arsifa et al. (2024) introduced the concept of language ego permeability, suggesting that learners who are more open to new linguistic and cultural elements may find it easier to accept and practice new pronunciation patterns.

Cognitive and metacognitive strategies also play an important role. Morley (1991) emphasizes self-monitoring, in which learners compare their own speech with a target model, and techniques such as shadowing support this process by requiring real-time processing and monitoring. Studies show that many learners use cognitive strategies such as memorizing phrases, focusing on pronunciation, and observing articulation, often supported by authentic materials (ERBAY* et al., 2016; Royani, 2023). Metacognitive strategies, including self-recording and consulting reference materials, further enhance learning, and explicit instruction that combines video viewing, repeated practice, and self-recording supports phonological awareness and pronunciation competence (Pardede, 2018).

Overall, pronunciation learning involves psychological factors, motivation and adaptability to the target language, interacting with cognitive and metacognitive strategies. Motivation plays a central role, particularly when reinforced through authentic materials, while systematic practice, self-monitoring, and explicit instruction provide the technical foundation for development (Pardede, 2018; Royani, 2023). These findings support integrated approaches that combine motivating authentic materials with structured cognitive training.

4. Teaching Pronunciation

Pronunciation teaching in EFL has evolved from emphasizing individual sound accuracy to prioritizing communicative intelligibility. Kenworthy (1987) defines this realistic goal as 'comfortable intelligibility' rather than native speaker accent, with more attention given to being understood in communication over phonetic perfection. Explicit pronunciation instruction proves more supportive than implicit approaches, especially when teachers develop students' phonological awareness (Yakut's 2020). This communicative philosophy provides the foundation for modern pronunciation instruction.

Supportive approaches teaching prioritizes aspects with greatest impact on intelligibility. Ur (1996) proposes a hierarchy: (1) word/sentence stress, (2) rhythm and intonation, (3) meaning-distinguishing consonants, and (4) vowel length distinctions. Recent research confirms that integrating segmental and suprasegmental features simultaneously produces superior results, particularly among Indonesian junior high school students (Al Mawardi et al., (2023). This integration simultaneously supports articulatory accuracy and builds students' oral communication confidence Gharamah et al. (2024).

Supportive pronunciation teaching requires multidimensional approaches. Kelly (2000) identified three essential components: imitative practice (listen-and-repeat), analytical awareness (stress marking, transcription), and communicative practice (role-play, games). Shadowing techniques exemplify this integration by requiring simultaneous attention to input and output Hamada (2018). Such explicit approaches combining video viewing, repeated practice, and self-recording facilitate phonological awareness and pronunciation competence Pardede (2018).

Integration with other language skills and affective factors proves essential. Pronunciation should connect with listening, speaking, reading, and writing activities to enhance transfer to real communication (Celce-Murcia et al. 2010). Video-based are useful for supporting pronunciation learning by combining visual and auditory elements, providing authentic language models while engaging students' motivation (Bajrami & Ismaili, 2016; Stempleski & Tomalin, 1990). Recent studies confirm that video-assisted pronunciation instruction facilitates students' pronunciation performance by enabling observation of articulatory movements while maintaining high engagement (Mukarrama & Fajriansyah, 2025). Balancing technical precision with learners' affective factors through such multimodal approaches proves crucial for supportive pronunciation instruction (Asrul & Husda, 2022).

Based on these perspectives, the present study adopts several guiding principles for implementing video-based shadowing in the classroom. First, pronunciation teaching is oriented toward communicative intelligibility rather than native-like accent, so the main goal is that students can be understood comfortably by listeners (Kenworthy, 1987). Second, instruction gives focused attention to key segmental contrasts that affect intelligibility, while also gradually raising learners' awareness of suprasegmental features such as stress, rhythm, and intonation (Al Mawardi et al., 2023; Celce-Murcia et al., 2010). Third, instead of conventional "repeat after me" drills, learners are engaged in real-time shadowing of video models, which combines imitative practice with continuous self-monitoring of pronunciation features (Hamada, 2018a; Kelly, 2000). Fourth, pronunciation work is closely linked to listening and speaking through authentic video that provide natural language models and rich contextual support (Bajrami & Ismaili, 2016; Stempleski & Tomalin, 1990). Finally, the teaching approach balances technical accuracy with learners' affective needs by

using engaging media and supportive procedures to build confidence, motivation, and reduced anxiety during pronunciation practice (Asrul & Husda, 2022).

Overall, pronunciation teaching is considered more balanced when it takes into account intelligibility, methodology, skill integration, and learner confidence. Approaches that prioritize key features, use varied techniques, and connect pronunciation with authentic communication and emotional support are often recommended. These principles support the use of innovative methods such as video-based shadowing in pronunciation instruction.

B. Video as Language Learning Media

According to Stempleski & Tomalin (1990), videos not only present sound but also visuals, sound effects, and music that provide contextual information about behavior, character, and culture which cannot be obtained from textual media alone. Videos are a distinct type of learning medium because they combine audio, visual images, and temporal sequencing simultaneously (Bajrami & Ismaili, 2016). Montero Perez (2022) emphasizes that watching audio-visual input integrates multiple sensory channels to enhance L2 comprehension and acquisition. Similarly, Alothali (2022) highlight how short videos deliver authentic multimodal experiences that combine audio, visual, and contextual cues to promote interactive language learning. Video can substantially increase students' motivation because it connects classroom learning with real-life communication and brings the outside world into the classroom (Bajrami & Ismaili, 2016; Elzeftawy, 2021). Language educators consistently report that authentic video helps learners experience language in naturally occurring situations rather than through artificial textbook examples, which encourages participation and more sustained practice (Bajrami & Ismaili, 2016; Polat & Erİştİ, 2019).

Polat and Erİştİ (2019) stated authentic videos are defined as materials created by native speakers for native audiences, rather than simplified products designed specifically for classroom teaching. In contemporary EFL teaching, video is no longer treated as an optional extra, but is widely recognized as a valuable medium for supporting English learning across proficiency levels (Alisoy, 2025; Bajrami & Ismaili, 2016). Authentic videos also expose learners to natural linguistic patterns and speech variation, including regional accents, reduced forms, and realistic prosody (Alisoy, 2025; Polat & Erİştİ, 2019). Because these materials are originally produced for native speakers, they create an "authentic linguistic environment" where learners encounter the full complexity of real language, such as fast speech rates and natural stress and intonation patterns (Polat & Erİştİ, 2019). One key advantage is that students can listen to unsimplified English as it is actually used by native speakers, which helps them develop more robust listening and pronunciation skills over time (Elzeftawy, 2021; Polat & Erİştİ, 2019). In addition, video supports comprehension because visual context clarifies meaning: images, facial expressions, and situational cues often convey meaning directly without the need for word-by-word explanation (Bajrami & Ismaili, 2016). When students repeatedly see authentic communication, cultural behavior, and language use integrated in real situations, their understanding becomes deeper and they gain more confidence to use English in real communication (Bajrami & Ismaili, 2016).

Recent studies show that video-based instruction is particularly valuable for supporting learners' pronunciation. In Indonesia, Mukarrama and Fajriansyah (2025) implemented an eight-week video-assisted pronunciation program with 32 senior high school students and found that mean pronunciation scores increased from 48.50 (poor) to 65.75, with a statistically significant gain ($t = -10.213$, $p = 0.000$). Internationally, Alisoy (2025) reported that six weeks of instruction with authentic video for 80 middle-school EFL learners led to clearer sound production, more accurate intonation, more precise word stress, and more natural speaking rhythm,

with learners also reporting high engagement and motivation. Stempleski and Tomalin (1990) emphasize that video is especially valuable for pronunciation because it allows students to see lip shapes, facial expressions, and tongue movements that match the sounds they hear, visual information that audio-only materials cannot provide.

To use video in a pedagogically sound way, teachers need to design lessons that actively engage students, carefully select appropriate videos, and structure viewing through three stages, pre-viewing, viewing, and post-viewing, combined with multiple exposures and focused tasks so that students can notice language patterns and consolidate what they have learned (Stempleski, 2018; Stempleski & Tomalin, 1990).

C. Shadowing Technique in Pronunciation Learning

1. Definition and Theoretical Basis of Shadowing

Shadowing is described as "a paced, auditory tracking task which involves the immediate vocalization of auditorily presented stimuli" (Lambert, 1992, p. 266). The shadowing technique is essentially a language learning method that requires learners to imitate the speech of native speakers simultaneously or immediately after hearing it. This activity, used initially in training simultaneous interpreters, fundamentally trains the cognitive capacity for shared attention between two complex tasks: listening to language input and producing language output simultaneously.

As it has evolved in the context of second language (L2) learning, the concept of shadowing has been enriched by researchers such as Hamada and Kadota. Hamada (2018a) distinguishes shadowing from conventional repetition techniques (repeat after me). According to him, shadowing is an online task that forces learners to focus their attention on the phonological (sound) aspects of the input received due to the very short time lag between hearing and speaking. In contrast, repetition is an offline task that allows for a longer pause, enabling

learners to process the meaning of the utterance. This fundamental difference makes shadowing a highly supportive tool for intensively training sound perception.

Therefore, the shadowing technique can be summarized as a language practice method by imitating what is listened to at almost or exactly the same time. Shadowing fundamentally trains the direct relationship between auditory perception (listening) and articulatory production (speaking), making it a very powerful tool for pronunciation development. Shadowing is not merely an imitation activity, but an intensive exercise that strengthens the connection between what is heard and what is spoken, and is often used to practice more accurate and automatic of pronunciation in a second language.

2. Cognitive Mechanisms in Shadowing

The advantage of shadowing in supporting pronunciation is rooted in how this technique works in the learner's cognitive system. Hamada (2018a), citing in Kadota (2007), explains that due to its demanding and high-speed nature, shadowing forces learners, especially those at the early to intermediate proficiency levels, to allocate their cognitive resources to phonological processing. In other words, this technique tends ‘blocks’ learners' access to deeper meaning processing, so that all attention is focused on capturing and imitating the details of the sounds heard.

Kadota (2019) explains how the brain works during the shadowing process. When students shadow, they cannot predict what will come next, so they must focus entirely on listening and repeating the sounds. Although this may seem like a limitation, it actually benefits learners. The brain's mechanism for temporarily storing speech sounds (phonological loop) has limited capacity, with information only lasting about two seconds before fading. By practicing shadowing repeatedly, learners train themselves to process

and retain pronunciation patterns more quickly and automatically, turning this limitation into an advantage in learning pronunciation.

In classroom practice, researchers recommend implementing shadowing through a series of scaffolded steps rather than as a single, one-off repetition task. Previous studies on shadowing for pronunciation propose procedures that begin with repeated listening to the target material, continue with silent shadowing to focus on articulatory movements, and then move to vocal shadowing with text support, sometimes followed by recording learners' performance for self-monitoring (Hamada, 2015, 2018a, 2018b; Kadota, 2019). Other work such as Murphey (2001) extends shadowing into more interactive formats, including conversational shadowing in pairs. Despite differences in format, these approaches share the same core principle: learners repeatedly imitate comprehensible input in real time in order to support the internalization of L2 pronunciation patterns.

Building on these principles, the present study adopted a step-by-step classroom procedure adapted to a junior high school EFL context and to the use of a short self-introduction video. The procedure included four key elements:

a. an audio familiarization phase in which students listened to the video without text;
b. shadowing practice at reduced speed without text;
c. scaffolded shadowing with text support and stress/intonation markings; and
d. integrative practice and communicative use, where students applied the pronunciation patterns in their own self-introductions.

These procedures provide gradual scaffolding from listening to full vocal production with self-monitoring, supporting the development of perception-production connections in pronunciation learning.

Both the cognitive explanation and these procedural principles provide a strong theoretical foundation for adapting shadowing in different learning contexts, including video-based shadowing with authentic materials for junior high school learners. Key ideas such as gradual scaffolding, selective focus on target features, and self-monitoring underpin the design of the activities used in this study.

3. The Benefits of Shadowing for Pronunciation

This process directly trains bottom-up processing skills, namely the ability to perceive sounds and recognize words based on their acoustic features (Hamada, 2018b). For many EFL learners, including secondary school students in Indonesia, the main weakness in listening and speaking skills often lies in bottom-up processing that has not yet been automated. By repeatedly training this aspect through shadowing, students can build a stronger phonological representation of English in their minds, which is an important foundation for pronunciation.

Theoretically, the shadowing mechanism, which directs students' focus to sound aspects of the input is closely related to pronunciation development. The process of imitating native speakers as accurately and as quickly as possible unconsciously trains learners to be more sensitive to detailed sound. According to Hamada (2018b), this includes segmental elements (phonemes such as vowels and consonants) and suprasegmental elements (word stress or emphasis, rhythm, and sentence intonation). When students actively try to replicate these elements, they not only train their ears to hear the differences, but also train their articulatory organs (tongue, lips, jaw) to produce these sounds.

The neurological basis of this dual training is the mirror neuron system, a network of neurons in the premotor cortex that is believed to underlie learning through imitation (Kadota, 2019). This system allows humans to learn through observation and imitation of speech sounds,

words, and phrases. Essentially, shadowing is not merely an artificial exercise, but rather activates the brain's natural mechanisms for language acquisition. When students shadow, they activate their innate neurobiological capacity that has evolved for language learning, making this repetitive exercise not only supportive but also fundamentally aligned with the way the human brain acquires language

Several studies in the Indonesian context have proven the that this technique can support pronunciation learning. A quasi-experimental study conducted by Sudrajat (2024) at SMAN 5 South Tangerang showed that the application of the shadowing technique significantly increased students' average pronunciation scores from 56.34 to 79.51, especially in terms of intonation and word stress. Similar results were found by Petalolo et al., (2024) in a study at SMP Negeri 3 Palu, where shadowing proved to be effective in improving the pronunciation of the interdental sounds /θ/ and /ð/, which are difficult for Indonesian learners, with a very significant increase in average scores. These findings indicate that shadowing can produce more natural-sounding speech and support pronunciation learning.

4. The Drawbacks of Shadowing for Pronunciation

Although shadowing has been described in the literature as beneficial, the technique presents a number of cognitive and pedagogical challenges that require special attention, especially when applied to junior high school students. One of the most critical drawbacks is the high cognitive demands that this technique places on students, who must listen, process, and reproduce speech in real time (Kadota, 2019). For beginner and intermediate EFL students, this multitasking demand can cause a heavy cognitive load, especially when the speech rate is too fast or the content is beyond their proficiency level. Research shows that when learners cannot keep up with the input, frustration and anxiety often arise, which can scrap the motivational

benefits intended by the shadowing technique (Matsui, 2011, in Kadota, 2019). Furthermore, the emphasis on phonological processing means that learners may engage in “meaningless repetition” without truly understanding the meaning of what they are saying (Hamada, 2014). This lack of semantic engagement limits the overall language learning value of shadowing and can result in students imitating sounds perfectly without being able to internalize vocabulary or recognize how words and phrases function communicatively.

The shadowing practice also poses inherent difficulties in self-monitoring and error correction. Because learners must maintain continuous, real-time vocalization to stay synchronized with the audio input, the cognitive space available to them to recognize and reflect on their own pronunciation errors is minimal (Kadota, 2019). Without explicit post-task reflection, such as listening to recordings of their own performance or receiving timely teacher feedback, students may inadvertently reinforce incorrect pronunciation patterns, causing errors to harden and become persistent features of their speech. Hamada's research specifically recommends that learners record and listen to their shadowing performances as a crucial metacognitive monitoring mechanism (Hamada, 2015). Furthermore, when shadowing is done without proper guidance or context, it risks becoming a purely mechanical exercise rather than a meaningful learning experience. Students may become skilled at imitating specific sounds from a particular recordings without developing phonological skills that can be applied to pronouncing new or unfamiliar words independently, especially if their practice is limited to a narrow range of structured or highly contextual audio material (Kadota, 2019; Lambert, 1992).

These challenges highlight that shadowing is not a “magic solution,” but a technique that depends heavily on careful instructional design, especially in the context of EFL classes. The way shadowing is implemented in class depends heavily on pedagogical mediation, such as the selection of materials appropriate to the students' level of proficiency. If the audio content is too difficult or has non-standard pronunciation, it can quickly become frustrating for students rather than motivating them (Hamada, 2011). Therefore, by recognizing and proactively addressing these potential obstacles through strategies such as procedural scaffolding, post-task reflection, and feedback integration, educators can mitigate risks and enhance the learning experience. Understanding the dual nature of shadowing, with its powerful benefits and inherent challenges, is precisely what makes exploring students' real experiences so important. This allows researchers to move beyond the question of “whether” this technique works, and instead ask “how” it works in real classrooms, taking into account all the complexities of the learning process.

5. Video-Based Shadowing Technique

Video-based shadowing is a specific form of shadowing that combines authentic video input with immediate oral imitation, allowing learners to copy native-speaker pronunciation in meaningful contexts (Martinsen et al., 2017; Mori, 2011; Van et al., 2024) Unlike audio textbook recordings, authentic video display various pronunciation patterns, varying phonetic phenomena, natural emphasis, and different speaking speeds, complexities difficult to display in simplified learning materials (Bajrami & Ismaili, 2016). Video-based shadowing is highly supportive in developing suprasegmental features such as stress, intonation patterns, and rhythm. Van et al. (2024) reported large gains in intonation, stress, and rhythm for Grade 10 students, and Alisoy (2025) found significant improvements in intonation accuracy, stress

recognition, and rhythm ability. When students learn from videos with strong visual support, suprasegmental features are easier to understand because students see head movements, facial expressions, and body language that accompany natural intonation patterns.

Video-based shadowing has advantages over audio-only shadowing because video provides visual information about how lips, tongue, and face move when speaking (Van et al., 2024). According to Stempleski and Tomalin (1990), video allows students to see actual communication and pause at certain moments to see details of lip and tongue positions. Alisoy (2025) confirms that watching native speakers helps learners master tongue/lip positioning through simultaneous visual, auditory, and kinesthetic input. This visual support proves especially valuable for consonant clusters and vowel distinctions, giving students clear articulatory models to imitate (Mukarrama & Fajriansyah, 2025). Martinsen et al. (2017) found high school learners improved read-aloud pronunciation through culturally contextualized video shadowing, valuing its authenticity. MICIK and RIZAOĞLU (2024) reported consistent gains in university EFL learners' comprehensibility, intonation, and speech rate. Van et al. (2024) documented large suprasegmental improvements in Grade 10 students. Locally, Tasyawfi and Lolita (2025) showed Indonesian seventh-graders using modified video shadowing outperformed conventionally taught peers on pronunciation post-tests, confirming effectiveness even for lower-proficiency junior high learners.

Although video-based shadowing is often regarded as beneficial, its implementation faces specific challenges that require appropriate pedagogical solutions. The main challenges include students having difficulty keeping up with fast speech rates of authentic videos, encountering unfamiliar vocabulary or cultural references that distract from pronunciation, and becoming overwhelmed by processing meaning, recognizing pronunciation patterns, and speaking

simultaneously (Alisoy, 2025; MICIK & RIZAOĞLU, 2024; Van et al., 2024). According to Krashen (1982), high anxiety can strengthen the affective filter. Therefore, the main solution is for teachers to carefully select videos with clear audio, moderate speech speed appropriate to students' level, relevant vocabulary, and extensive previewing activities to reduce anxiety and build background knowledge (Stempleski & Tomalin, 1990). In addition, teachers are encouraged to structure video-based lessons in clear stages by carefully choosing video length (3–5 minutes), planning multiple viewing cycles with different focus objectives, and providing strategic corrective feedback that is positively framed to maintain motivation while addressing specific errors(Krashen, 1982; Stempleski, 2018).

D. Study of Relevant Research

A number of studies report positive roles of shadowing in pronunciation development, including studies with Indonesian junior high school students.. Petalolo et al. (2024) conducted quasi-experimental research at SMP Negeri 3 Palu with seventh-graders, showing dramatic improvements in experimental groups from 38.9 (very poor) to 80.8 (very good), confirming shadowing's effectiveness at the secondary level in Indonesia. This quantitative evidence is strengthened by Hamada (2018b), a leading shadowing researcher, who tested two advanced variations haptic-shadowing and IPA-shadowing and found both significantly improved comprehensibility, segmental accuracy, and suprasegmental accuracy. Together, these studies establish shadowing's strong theoretical and practical foundation.

Qualitative perspectives reinforce these quantitative findings. Pratama and Isnaini (2024) revealed through case study interviews that students perceived shadowing with Text-to-Speech AI as flexible and effective, reporting increased phonological awareness and intonation understanding alongside enhanced speaking confidence. Similarly, Utami &

Morganna (2022) conducted classroom action research with ninth-graders at SMPN 1 Curup Timur, documenting significant pronunciation improvements through shadowing.

Video-based shadowing is a specific form of shadowing that combines authentic video input with immediate oral imitation. Martinsen et al. (2017) showed that high school learners improved pronunciation and valued the authenticity and autonomy of video-based shadowing and tracking. MICIK and RIZAOĞLU (2024) reported small but consistent gains in comprehensibility, intonation, and speech rate for university EFL learners, with generally positive attitudes. Van et al. (2024) found a large effect on intonation for Grade 10 students (effect size = 1.50), with almost all learners expressing positive views of the technique. Tasyawfi and Lolita (2025) showed higher post-test pronunciation scores for Indonesian seventh-graders taught with modified video-based shadowing than for those taught conventionally, indicating effectiveness across secondary and tertiary levels.