CHAPTER I
INTRODUCTION

This chapter presents the research background, research questions, research purposes, research significances, research scope, conceptual framework, hypothesis, and previous studies.

A. Research Background

Listening comprehension in English as a Foreign Language (EFL) involves more than knowing vocabulary or grammar. Meng et al. (2023) It is described as an active cognitive process in which listeners decode sounds, analyse syntactic structures, and extract meaning from continuous speech. An important yet often unnoticed aspect of this process is prosody, the suprasegmental features of spoken language that include stress, intonation, rhythm, pitch changes, and pauses. Gussenhhoven & Chen (2020) Describe prosody as the structural pattern of spoken discourse that indicates word boundaries, emphasises information focus, reflects syntactic connections, and conveys pragmatic meanings beyond the literal words. In research on L2 listening, prosodic cues include acoustic signals such as stressed syllables, pitch variations, and rhythmic groupings, which listeners rely on to analyze and understand spoken language in real time (Yenkimaleki et al., 2023)

The role of prosodic cues in EFL listening comprehension is well documented. Meng et al. (2023) and Yenkimaleki et al. (2023) argue that prosodic signals serve as bottom-up scaffolds during speech processing, guiding listeners toward accurate segmentation of continuous speech and supporting higher-level inferential and pragmatic understanding. When prosodic cues are misread or ignored, comprehension at all three levels (literal, inferential, and pragmatic) may break down, even when the listener possesses sufficient vocabulary and grammatical knowledge. This is particularly relevant in EFL contexts, where learners' first-language (L1) prosodic systems may differ substantially from English, creating a challenge

of prosodic transfer that affects the perception of emphasis, sentence modality, and speaker attitude (Nagle & Baese-berk, 2022).

In the Indonesian EFL context, this prosodic challenge is commonly documented. Indonesian is a syllable-timed language where syllables have roughly equal duration and weight, unlike English, which is stress-timed with stressed syllables contrasting sharply with reduced ones. This typological difference can make it difficult for Indonesian EFL learners to grasp English prosodic cues, including lexical stress and intonation contours, which are essential for parsing and understanding meaning (Nagle & Baese-Berk, 2021). Harianja et al. (2022) noted that Indonesian senior high school students identified rapid speech and unfamiliar intonation as major barriers to listening comprehension, indicating a prosodic issue rather than purely lexical or grammatical ones. The study also revealed that these students consistently perform poorly on listening tasks involving naturalistic spoken English with authentic prosodic features. Since the national curriculum (Kurikulum Merdeka) expects students to demonstrate communicative competence in real listening situations, this highlights a significant gap between curriculum goals and actual learner performance.

A growing pedagogical approach involves using short video-based media that showcase natural prosodic features in multimodal settings. Segments lasting 15 to 60 seconds, similar to TikTok, Instagram Reels, and YouTube Shorts, display prosodic cues both auditorily and visually through gestures, facial expressions, and physical emphasis that correspond with stress and intonation. Toyama & Hori (2025) suggest that multimodal input may enhance L2 prosodic perception, as visual cues such as pitch and beat gestures offer learners additional channels for interpreting suprasegmental features, which can be hard to detect from audio alone. Zhang et al. (2023) also observed that multimodal cues, including prosody and gestures, affected how L2 learners processed spoken words. Nonetheless, L2 listeners benefited less than L1 listeners, implying that explicit focus on prosodic features in such input could support their role in understanding. The short

length of these videos might also reduce the cognitive load of processing natural speech, helping learners concentrate more effectively on specific prosodic features within smaller input segments.

The relevance of short videos as an instructional medium is particularly salient in the Indonesian adolescent context. According to *Asosiasi Penyelenggara Jasa Internet Indonesia* (APJII) (2024) 87% of Generation Z users in Indonesia, including current high school students, are active daily internet users, and short-video platforms are the dominant form of daily media consumption. Kemp, (2025) reports that Indonesian users spend an average of over 38 hours per month on TikTok, making it the most time-intensive digital platform in the country. Purboningsih et al. (2023) In a Bandung-based study of Indonesian adolescents aged 12–15, researchers found that platforms such as TikTok and YouTube are embedded in students' everyday routines for entertainment and informal learning. This media familiarity may lower the affective barrier to engaging with short-video content in learning contexts and create conditions for more motivated and sustained exposure to English prosodic input.

Although there is evidence supporting the use of multimodal videos, and short videos are popular among Indonesian adolescents, questions remain about how prosodic cues in short videos affect Indonesian EFL high school students' listening comprehension. Most research on prosodic instruction uses traditional materials or controlled recordings (Yenkimaleki et al., 2023), and studies on short videos in Indonesian EFL settings tend to focus more on motivation and exposure rather than prosody (Setiawan et al., 2024). The connection between prosodic cues in short videos and EFL listening covering literal, inferential, and pragmatic dimensions is limited, especially in Indonesian senior high schools. This study explores how prosodic cues in short videos affect Grade 11 students' listening comprehension.

B. Research Questions

Based on the background problem above, the research question is formulated as follows:

1. What are EFL students' listening comprehension skills taught with prosodic cues in short videos?
2. What are EFL students' listening comprehension skills taught without prosodic cues in short videos?
3. Is there any significant difference in EFL listening comprehension skills between students who receive prosodic cues in short videos and those who do not?

C. Research Purposes

Based on the research questions above, the goals of this study are as follows:

1. To examine the listening comprehension skills of EFL students who are taught using prosodic cues in short videos.
2. To examine the listening comprehension skills of EFL students who are taught without prosodic cues in short videos.
3. To find out whether there is a significant difference in EFL students' listening comprehension skills between those who receive prosodic cues in short videos and those who do not.

D. Research Significances

This study offers both theoretical and practical significance in EFL listening instruction, especially in Indonesian secondary education. By analyzing how prosodic cues in short videos affect listening comprehension, this study addresses a pedagogical concern related to phonological processing, technology-enhanced language learning, and Indonesian adolescents' digital media habits. The findings may interest researchers in prosody and L2 listening, as well as teachers, curriculum developers, and school decision-makers seeking evidence-based approaches.

Theoretically, this study contributes to the literature by applying Anne Cutler's Metrical Segmentation Strategy (MSS) to examine how prosodic cues in short videos influence EFL listening comprehension. MSS suggests that listeners use prosodic cues, especially lexical stress, to segment speech in real time. Originally developed for first-language contexts, MSS has limited applicability to second-language learners, especially those from syllable-timed backgrounds such as Indonesian. This study extends MSS into a multimodal, short-video environment in which prosodic cues are conveyed through audio and visual signals, such as gestures and facial expressions. It aims to deepen understanding of how naturalistic, interest-driven videos support speech segmentation and comprehension across literal, inferential, and pragmatic levels. The findings could inform whether prosodic segmentation operates differently in brief, visually enriched digital media compared with other listening activities.

Practically, this study offers Indonesian EFL teachers a low-cost, relevant approach by using short videos familiar to adolescents as a medium for exposing students to natural prosodic cues. Instead of costly listening materials, teachers can use free videos from platforms like YouTube Shorts or TikTok that are clear, have varied intonation, and maintain a natural rhythm. This is especially helpful in resource-limited schools where authentic listening input is scarce. Regular exposure to prosodic cues in short, visually supported videos may reduce listening anxiety by making natural speech more predictable and engaging for students. At the institutional level, findings can guide decision-makers and curriculum developers in integrating short-form digital media into listening instruction, aligning with the national curriculum standards (Kurikulum Merdeka).

E. Research Scope

This study investigates the effect of prosodic cues, specifically word stress and intonation, delivered through teacher-selected short videos, on the EFL listening comprehension of Grade 11 students at a single public high school in Bandung, Indonesia. Listening comprehension is assessed across

three dimensions, including literal, inferential, and pragmatic comprehension, measured through a pre-test and post-test instrument. The study follows a quasi-experimental design conducted in three treatment sessions, excluding the pre-test and post-test administration. The instructional media is limited to short videos of 15 to 60 seconds in duration featuring clear, near-native English speech, selected by the teacher for prosodic clarity and content appropriateness. The study involves two intact classes from a single school, with one serving as the experimental group and the other as the control group. The scope of this study is bounded as follows. It focuses on word stress and intonation as the targeted prosodic cues, excluding rhythm, pausing, and pitch variation as separate variables. It measures comprehension outcomes only at the time of the post-test and does not examine long-term retention. It does not address pronunciation development, vocabulary acquisition, speaking skills, or the use of other digital platforms beyond teacher-selected short videos.

F. Conceptual Framework

This study is based on the idea that EFL listening comprehension relies not just on vocabulary and grammar skills but also on learners' ability to interpret prosodic cues in spoken language. As shown in the conceptual framework above, prosodic cues in short videos are treated as the independent variable, while EFL listening comprehension is the dependent variable, assessed through pre-test and post-test measurements.

The theoretical foundation for the independent variable is based on Anne Cutler's Metrical Segmentation Strategy (MSS, 1990). This theory suggests that prosodic cues, especially word stress and intonation, are the primary acoustic signals listeners use to segment continuous speech into meaningful units during real-time comprehension. In this framework, prosodic features are not merely secondary aspects of spoken language but essential signals that mark word boundaries, highlight information focus, and indicate syntactic and pragmatic structures. Therefore, the study's instructional focus centres on two specific prosodic cues, word stress and

intonation. These were chosen because they are the most acoustically prominent and pedagogically accessible suprasegmental features for EFL learners at the senior high school level. These cues are conveyed through teacher-selected short videos lasting 15 to 60 seconds, providing naturalistic, multimodal input in which prosodic signals are reinforced both auditorily and visually through gestures, facial expressions, and physical emphasis.

The dependent variable, EFL listening comprehension, is defined and assessed based on the framework by Vandergrift and Goh (2021). They describe listening comprehension as a layered process involving both bottom-up decoding and top-down meaning-making. According to this framework, listening comprehension comprises three key aspects. Literal comprehension measures the ability to recall explicitly stated information. Inferential comprehension involves drawing conclusions from implied or incomplete data. Pragmatic comprehension involves understanding the speaker's intent, attitude, and contextual meaning beyond the literal words. These three aspects are evaluated through a listening test conducted as both a pre-test and a post-test.

The connection between the two variables is derived from the segmentation function described by MSS. When learners correctly identify word stress and intonation from short videos, their speech parsing improves, leading to better comprehension at literal, inferential, and pragmatic levels. The brief length of these videos aims to reduce cognitive load, enabling learners to focus more on meaning construction rather than handling excessive input. Accordingly, the framework suggests that students in the experimental group, who are exposed to prosodic cues through short videos, are likely to show notable improvements in listening comprehension across all three areas compared with the control group, which learns listening without a focus on prosody.

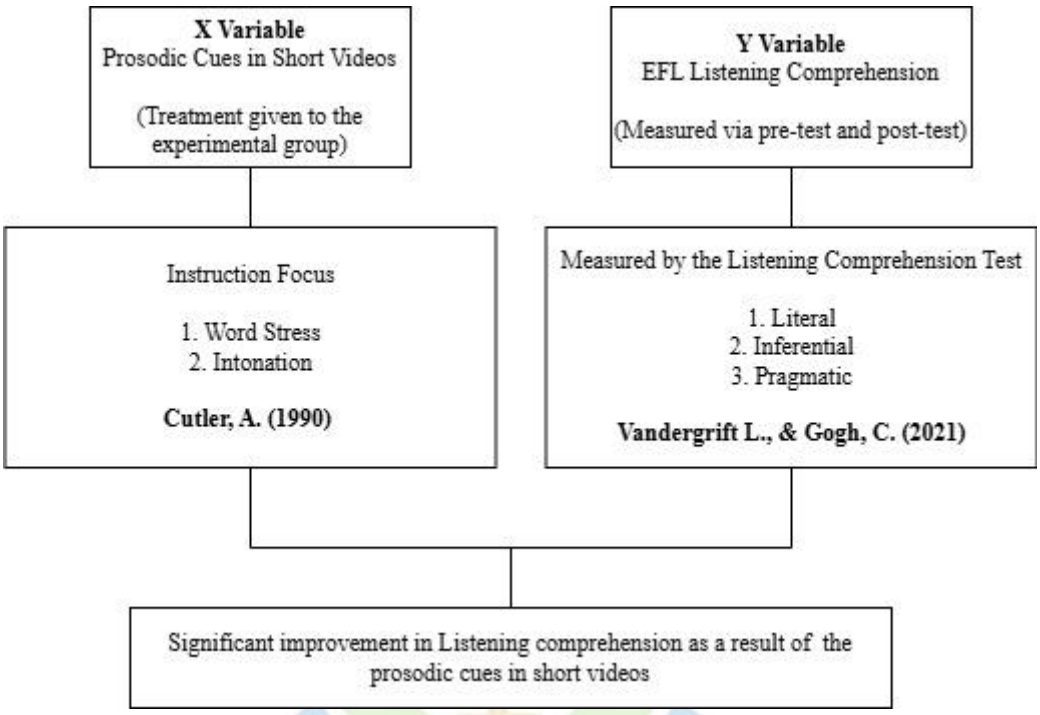

Figure 1. 1 Conceptual Framework

G. Hypothesis

This study's hypotheses are grounded in theoretical and empirical foundations that highlight the role of prosodic cues in facilitating listening comprehension. Drawing on Cutler's (1990) Metrical Segmentation Strategy (MSS), prosodic cues, particularly word stress and intonation, function as primary acoustic signals that listeners use to segment continuous speech into meaningful units during real-time processing. This segmentation function is considered fundamental to comprehension, as it precedes and enables higher-order processes such as inferential and pragmatic interpretation. Supporting this theoretical basis, Yenkimaleki et al. (2023) provide empirical evidence that exposure to prosodic cues in instructional settings may improve EFL learners' ability to decode spoken input and construct meaning across multiple levels of comprehension. Within a multimodal short-video environment, these cues may be further reinforced by visual signals, such as speaker gestures and facial expressions, which Zhang et al. (2023) suggest can reduce processing difficulty for L2 listeners. On this basis, the following hypotheses are proposed:

1. Null Hypothesis (H_0): There is no significant difference in the listening comprehension skills of Indonesian EFL high school students who receive prosodic cues in short videos and those who receive regular listening lessons without a focus on prosodic cues.
2. Alternative Hypothesis (H_1): There is a notable difference in the listening comprehension skills of Indonesian EFL high school students who receive prosodic cues in short videos compared to those who receive standard listening lessons without a focus on prosodic cues.

Although the statistical hypothesis is non-directional, this study predicts that students exposed to prosodic cues in short videos will achieve higher post-test listening comprehension scores across literal, inferential, and pragmatic aspects than those not exposed to such cues.

H. Previous Studies

A significant body of research emphasizes the importance of prosodic features, especially stress and intonation, in enhancing listening comprehension among EFL learners. Yenkimaleki et al. (2023) conducted an experimental study showing that explicit teaching of suprasegmental features, such as word stress and intonation, led to notable improvements in listening scores compared to a control group without prosody-focused instruction. Their results indicate that highlighting prosodic cues during lessons can boost bottom-up decoding and overall speech understanding. Nonetheless, since their study used traditional classroom and audio materials, it remains unclear whether similar results can be achieved with prosodic cues presented in short-form digital videos rather than conventional teaching media.

A recent meta-analysis by Ngo et al., (2024), which synthesised 15 experimental and quasi-experimental studies on technology-assisted pronunciation training, found that such interventions have only a small effect on suprasegmental features such as stress and intonation, in contrast to a considerably larger effect on segmental (individual sound) features. Critically, their moderator analysis revealed that short treatment durations

of one to four weeks yielded a negligible, statistically non-significant effect, compared with medium- or long-duration interventions. This finding suggests that the efficiency of suprasegmental instruction may be highly sensitive to treatment duration, raising the question of whether brief instructional interventions, such as those feasible within a limited classroom schedule, are sufficient to produce measurable listening-comprehension gains among EFL learners.

Investigating a closely related short-video format, Dukhayel, (2021) examined the effect of different caption conditions on 96 EFL learners' comprehension of vlogs, a form of short, informal video content structurally similar to the short videos examined in the present study. The results showed that the availability of captions did not consistently produce significantly better listening comprehension, particularly among lower-proficiency learners, who were unable to simultaneously process the multiple modalities of images, audio, and text within their limited working-memory capacity. This finding illustrates that multimodal short-video input does not automatically translate into significant comprehension gains and may instead impose a cognitive-processing cost that offsets its potential benefits, a tension directly relevant to the design of prosodic-cue instruction delivered through short videos.

Research in Indonesia's EFL context points to specific challenges students face with prosodic features during listening activities. Harianja et al., (2022) found that rapid speech and unfamiliar intonation patterns are the main listening barriers reported by senior high school students in Pematangsiantar. Their survey revealed that prosodic elements, rather than vocabulary limitations alone, are the key difficulties in understanding, highlighting the importance of focusing on prosodic cues in listening lessons for Indonesian secondary schools. However, since the study was purely descriptive and did not evaluate any prosodic teaching methods, it left open the question of how effective such instruction might be.

Building on this, Rizky & Setiawan, (2024) investigated the effectiveness of YouTube Shorts in enhancing listening comprehension among senior high school students at SMAN 1 Kademangan Blitar. Their results showed a notable increase in post-test scores after using the videos, indicating that short-form videos could be a practical and accessible tool for advancing EFL listening skills in Indonesian secondary schools. Nonetheless, they did not analyze the prosodic features in the videos nor differentiate among comprehension levels such as literal, inferential, or pragmatic understanding, leaving the specific role of prosodic cues in this context unexplored.

Complementing this, Setiawan et al. (2024) found that undergraduate EFL students in Indonesia viewed short-form platforms such as TikTok positively as tools for developing listening and speaking skills. Students reported that exposure to natural speech in short videos increased their awareness of pronunciation and prosodic variation. However, the study also noted that the fast speech rate and informal prosodic patterns common in short videos posed comprehension challenges, particularly for learners who lacked prior guidance on interpreting prosodic cues. These findings suggest that while short videos may increase exposure to prosodic features in English, pedagogically structured engagement with those features may be necessary to translate that exposure into comprehension gains.

Overall, research shows that prosodic cues strongly affect EFL listening and suggests that short videos are effective for prosodic input in Indonesian EFL. Limited data exist on how features such as stress and intonation in short videos affect listening among Indonesian high school students. This study addresses this gap by presenting results from a quasi-experimental study in Indonesian schools.