CHAPTER I
INTRODUCTION

A. Research Background

The world has evolved at a fast rate, and technology has significantly impacted the process of teaching and learning. In the English language classroom, technology offers audio-visual aids that allow students an alternative learning experience that can replace the use of the textbook. The use of audio-visual materials as language presentation tools, in the context of theme-based language learning, such as movies (Bajrami & Vela, 2017), supports the presentation of the language, which is learner centered and focuses on learner autonomy, and enriches the process of interaction in the classroom. The use of the audio-visual media (movie clips) has proven to be relevant mode in English language learning especially for EFL students.

This technological shift is particularly crucial when addressing the current demographic of learners, widely recognized as Generation Z or "digital natives" (Prensky, 2001). Today's seventh-graders are living in a hyper-connected, multimedia-rich world where watching videos, playing video games, and using social media are a way of life. This means that traditional, monolithic teaching approaches like memorizing from textbooks or listening to a tape recorder without seeing the person speaking can lead to high levels of cognitive disengagement and low levels of intrinsic motivation. Educators need to make a connection between students' out-of-school digital lives and the formal pedagogy in school to effectively engage students and support meaningful second language acquisition. The use of authentic movie clips makes the learning context more than a mere academic task, it becomes a highly contextualized linguistic experience. The affective filter is lowered by using the media format that students choose in their daily lives, and the cognitive readiness to absorb complex phonological features is greatly increased.

English as a foreign language (EFL) students in Indonesia still have great difficulty in mastering the phonological level in their English language learning, although technology is being applied in the EFL teaching field. One of the most difficult skills for EFL learners is pronunciation. Often, there is a difference between the phonetic system and sequences of sounds in the learner's L1 and L2. Thus, many students have problems

UNIVERSITAS ISLAM NEGERI
BANDUNG

in expressing both segmental features (accuracy of vowels and consonants) and suprasegmental features (word stress and intonation). Students' ability to produce these elements is difficult if they lack sufficient and suitable exposure to spoken English in the real world.

Movie clips are extensively used in order to compensate for this lack of exposure, to improve pronunciation. According to the theory of the Multimodal Learning of Mayer (2009), the information that is processed by the learner is better when presented in the form of visual and verbal. The movie clip exposure (X_1) promotes this end by providing real spoken English and visual feedback. The exposure in the present study is not only the number of movies watched, but also the level of involvement of the learners in watching movies, such as watching the mouth movements of the actors to be able to produce accurate sounds, watching the actors' face movements to understand their intonation and emotion.

Subtitle awareness (X_2) is also an important cognitive component that assists pronunciation development, in addition to audio-visual input. English also has irregular sound-to-spelling correspondences, and this can lead to students not being able to pronounce words correctly when reading them in isolation or when reading them without the support of a text. Subtitle awareness is the learners' mental activity to use on-screen texts to confirm spoken information and to correspond sounds with their correct spelling.There are studies that indicate that subtitles enable students to concentrate on the pronunciation part of the language, including its sounds, stress and word forms (Mishbahuddin, 2025). With this approach, students who have high awareness of subtitles will be able to create good sound spelling mapping when conducting learning activities by watching movies.

Although a number of studies have focused on the use of movies and subtitles in EFL learning, existing literature has predominantly focused on learners' perceptions (Damayanti et al., 2025; Putri & Zuraida, 2025; Salsabila, 2022), general vocabulary acquisition (Baranowska, 2020), or broad speaking fluency without phonetic isolation (Laoli et al., 2025). Little research has been done to test the effect of exposure to movie clips and awareness of the subtitles as measurable, independent psychological constructs on the pronunciation accuracy of L2 learners. In addition, there is a clear methodological

gap in the assessment of these variables. This study is a Multiple Linear Regression approach, as opposed to experimental designs which only assess pre- and post-test differences. This specific statistical design is employed because it allows the researcher to quantify the exact predictive relationship and the percentage of variance in pronunciation proficiency that is jointly and individually contributed by the two independent variables without manipulating the natural learning environment (Creswell, 2014; Field, 2018).

Therefore, this study examines the correlation between the exposure to movie clips, subtitle awareness, and English pronunciation among seventh-graders. Finally, the answer to the above gaps, the end result of this research offers empirical evidence in the form of concrete statistics that passive exposure to movie clips alone is not enough for phonetic acquisition, but when combined with high subtitle awareness, it plays a significant role in students' phonological accuracy. This confirms the absolute need for textual cognitive scaffolding in using bimodal media in EFL classrooms.

B. Research Questions

Based on the background outlined above and the multiple linear regression design applied in this study, the research focuses on the following problems:

1. Is there a significant partial correlation between movie clip exposure (X_1) and seventh-grade students' English pronunciation skills (Y)?
2. Is there a significant partial correlation between subtitle awareness (X_2) and seventh-grade students' English pronunciation skills (Y)?
3. Is there a significant simultaneous correlation between movie clip exposure (X_1) and subtitle awareness (X_2) on seventh-grade students' English pronunciation skills (Y)?

C. Research Purposes

In relation to the specific research questions, the purposes of this research are:

1. To determine the significance of the partial correlation between movie clip exposure (X_1) and seventh-grade students' English pronunciation skills (Y).

2. To determine the significance of the partial correlation between subtitle awareness (X_2) and seventh-grade students' English pronunciation skills (Y).
3. To determine the significance of the simultaneous correlation of both movie clip exposure (X_1) and subtitle awareness (X_2) on seventh-grade students' English pronunciation skills (Y).

D. Research Significances

This study is likely to make significant contributions in several areas:

1. Theoretical Significance: The results of this study have theoretical implications in developing theoretical perspectives in pronunciation learning in EFL using audio-visual media. It empirically validates Mayer's Multimedia Learning Theory by showing how segregating the visual-auditory input (movie exposure) and textual input (subtitle awareness) provides a cognitive contribution to the acquisition of English phonology.
2. The contributions of this study to practical application for the English teacher are that they can use the findings from the study to design more effective pronunciation learning activity by guiding students' attention in order to utilize visual learning and subtitles. It is a sort of conscious map of English sounds to spellings for students to use independently of their movies with subtitles.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

E. Research Scope

Research Scope and Limitations In order to keep the study focused and scientific and in accordance with the objectives of the research, the researcher sets the following boundaries:

1. The independent variables include only two: X_1 (as measured by visual and auditory engagement of the movie clips) and X_2 (as measured by the cognitive use of the subtitles in the movie clip, using Likert-scale questionnaires).
2. Skill Limitation and Scoring System: only English Pronunciation Skills are considered dependent variable (Y). This study does not consider other aspects of speech that include fluency, grammar or vocabulary breadth. Only three specific

phonological indicators are focused on in the evaluation: Accuracy of target sounds, Word stress and Intonation. For objectivity, a binary scoring system (Hit or Miss) is used for 15 predetermined movie dialogue sentences and 18 predetermined target words, which are multisyllabic words (with 48 max raw scores).

3. Participant Limitation: This study was carried out in the academic year 2025/2026 using the total sampling technique in junior high school with a class of 27 Grade 7 students. The results were therefore be specific to the group of children studied and might not be fully generalizable to other children.

F. Conceptual Framework

The conceptual framework of this study is grounded in the Cognitive Theory of Multimedia Learning (Mayer, 2009), which emphasizes that optimal language acquisition occurs when learners actively process information through both visual and verbal channels. Since this study employs a multiple linear regression approach, this framework aims to map the independent variables' partial and simultaneous predictive correlation with the dependent variable, rather than establishing direct experimental causality.

The relationships between the variables in this study are visualized in the following diagram:

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

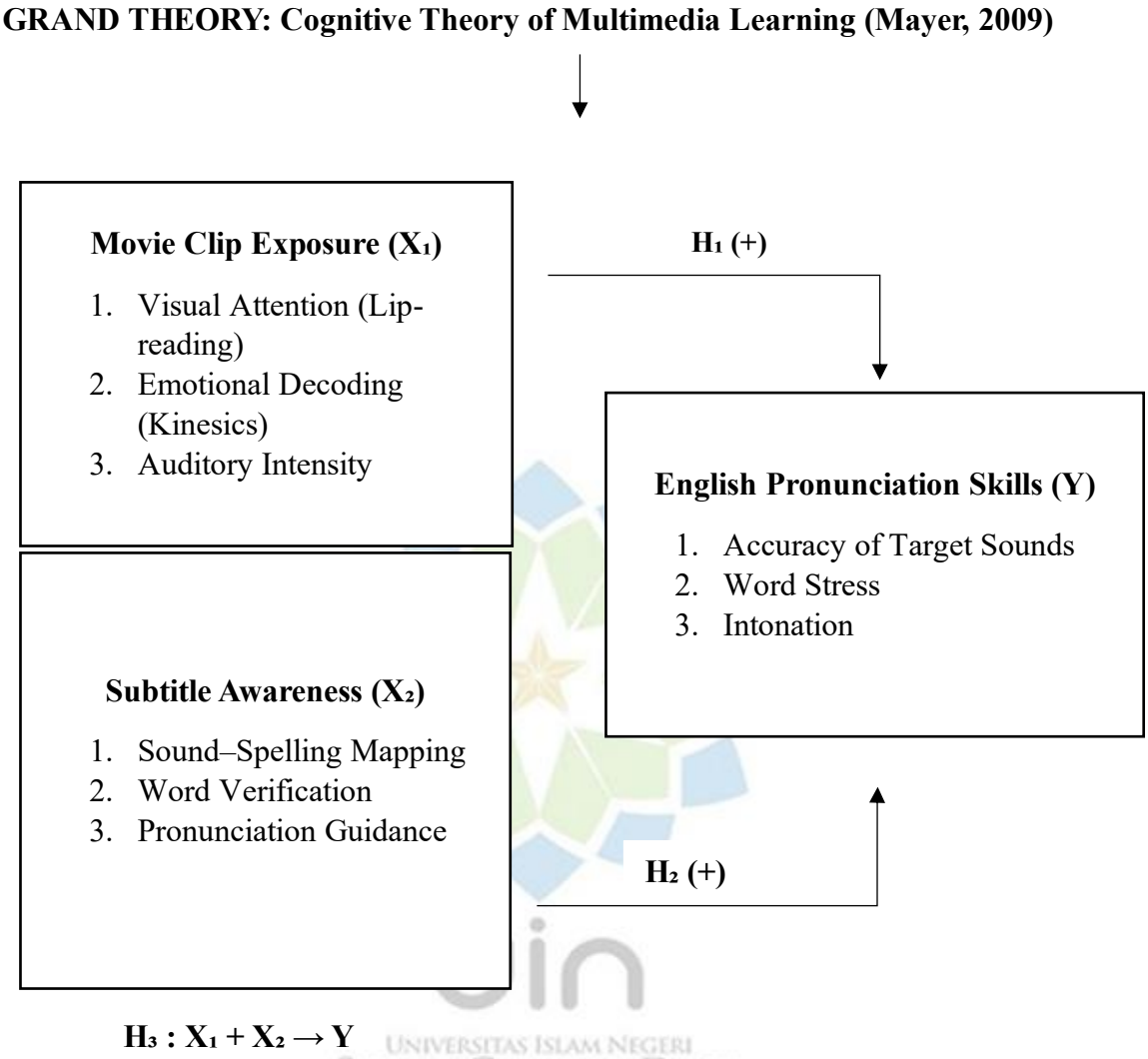


Figure 1. 1 *Conceptual Framework Diagram*

Note: *(H_1) and (H_2) represent partial correlations, while (H_3) represents simultaneous correlation.*

Based on the diagram above, the conceptual framework operates as follows:

1. **Movie Clip Exposure (X_1):** This variable acts as the primary authentic bimodal input. It provides real spoken English accompanied by vital visual feedback (facial expressions and mouth movements). Active exposure allows learners to observe the physiological articulation of sounds and decode the emotions that govern natural intonation patterns.

2. **Subtitle Awareness (X_2):** This variable acts as the cognitive scaffolding. Given the deep orthography of the English language, learners utilize written text (subtitles) to perform sound-spelling mapping. This cognitive awareness helps them recognize word boundaries and verify the exact syllables that carry primary stress.

3. **English Pronunciation Skills (Y):** This is the output variable. It reflects the learners' psychomotor ability to accurately produce segmental sounds (accuracy) and suprasegmental features (word stress and intonation) as a direct result of processing the aforementioned inputs.

Comprehensively, this framework hypothesizes that when high exposure to visual-auditory cues is synergized with a high cognitive awareness of subtitles, they will positively and significantly correlate with the students' English pronunciation performance.

G. Hypothesis

The present study is a quantitative correlational research design with Multiple Linear Regression to study the statistical and predictive relationship between two independent variables and one dependent variable without intervening the natural learning environment. The exposure to movie clips is considered as variable X_1, the awareness of the subtitles as variable X_2, and the English pronunciation skills as variable Y. The analysis is intended to determine the existence, strength, direction, and statistical significance of the variables' relationships, both partially and simultaneously.

In a correlational study, a directional hypothesis can be formulated if theoretical frameworks and prior empirical research strongly suggest a specific direction of the relationship. Drawing on Mayer's (2009) Cognitive Theory of Multimedia Learning and earlier studies reviewed in this research, it is anticipated that both bimodal inputs will positively correlate with pronunciation accuracy. Accordingly, the hypotheses of this study are divided into partial and simultaneous formulations, stated as follows:

1. First Hypothesis (Partial Correlation of X_1)

a. Alternative Hypothesis ($H\alpha_1$): There is a significant positive partial correlation between X_1 and Y.

b. Null Hypothesis (H_{01}): There is no significant positive partial correlation between X_1 and Y.

2. Second Hypothesis (Partial Correlation of X_2)

a. Alternative Hypothesis ($H\alpha_2$): There is a significant positive partial correlation between X_2 and Y.

b. Null Hypothesis (H_{02}): There is no significant positive partial correlation between X_2 and Y.

3. Third Hypothesis (Simultaneous Correlation of X_1 and X_2)

a. Alternative Hypothesis ($H\alpha_3$): There is a significant positive simultaneous correlation between X_1 and X_2 on Y Variable.

b. Null Hypothesis (H_{03}): There is no significant positive simultaneous correlation between X_1 and X_2 on Y Variable.

These hypotheses will be tested at a standard significance level of $\alpha = 0.05$. The partial correlations (H_1 and H_2) will be evaluated using the t-test statistic to see the individual contribution of each variable. Meanwhile, the simultaneous correlation (H_3) will be evaluated using the F-test statistic in the ANOVA table. This robust approach determines exactly how the variables interact to predict EFL students' phonological acquisition.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

H. Previous Studies

There has been a significant amount of literature that has examined the impact of multimedia learning, digital streaming, and the use of subtitles on English proficiency, especially in the context of English as a Foreign Language (EFL). Such as the studies are important theoretical and empirical foundations for the current research, and they also indicate the specific methodological research gaps that will be addressed by the current research, which is based on a rigorous statistical design.

In their seminal study, Bajrami and Vela (2017) examined the general effectiveness of audio-visual media, such as movie clips, in language learning. They found that the use of audio-visual tools in EFL classes helps to create active interaction and student-centered learning, which makes students active observers instead of passive receivers. They found that multimodal input enhances learners' contextual understanding. Extending this line of

research, Laoli et al. (2025) explored the role of English movie viewing in supporting students' speaking skills. They found that watching English movies serves as a critical factor in building vocabulary and speaking confidence because the visual and aural stimuli act as authentic models for learners to imitate. While these studies confirm the foundational pedagogical benefits of multimedia, they treated language proficiency broadly and did not isolate the specific, granular pronunciation components (segmental and suprasegmental features).

Addressing the specific physical aspect of audio-visual observation, recent studies highlight the role of visual articulatory cues in L2 phonological acquisition. Li and Hua (2014) examined the effectiveness of visual cues in assisting Chinese EFL learners to distinguish non-native consonant contrasts, revealing that visual information significantly enhances phonetic perception when it coincides with audio information. Furthermore, Toyama and Hori (2025) emphasized that multimodal approaches involving speech visualization and facial cues provide intuitive guidance for learners to internalize challenging L2 pronunciation. This supports the fundamental rationale for the inclusion of Movie Clip Exposure in the present research, positioning the act of watching not merely as entertainment, but as a crucial physiological guide for articulatory placement.

Advancing the discourse on digital learning and textual accessibility, Putri and Zuraida (2025) conducted a comprehensive mixed-method study on the perceptions of university students using English subtitles on streaming platforms like Netflix. Grounded in the Technology Acceptance Model (TAM), they discovered that learners heavily rely on subtitles as an intentional strategy for vocabulary enrichment, listening comprehension, and notably, pronunciation and grammar awareness. Their research reinforces the widespread acceptance of this bimodal method in the Indonesian educational landscape. However, their study was largely perception-based. The academic debate surrounding cognitive overload and on-screen text was rigorously examined by Baranowska (2020). In her experimental study on EFL learners, she tested the effects of different subtitling conditions on cognitive load and comprehension. Counterintuitively, her findings demonstrated that viewing English videos with English subtitles actually lowered the learners' cognitive load compared to viewing without subtitles, while simultaneously increasing comprehension. Baranowska (2020) concluded that subtitles do not result in cognitive overload; instead, they free up mental resources for learning. This is further

supported by Kruger and Doherty (2016), who argued that same-language subtitles can be used to manage cognitive load efficiently, particularly for students learning in a second language.

This cognitive approach is strongly supported by the experimental findings of Mishbahuddin (2025), who investigated the specific effects of videos with subtitles on pronunciation ability among EFL learners. The study reported a dramatic improvement in passing rates from 35% to 90%, confirming that bimodal input the simultaneous intake of audio and text is particularly helpful for learners who have severe difficulties differentiating between nuanced pronunciation aspects. Furthermore, Salsabila (2022) and Damayanti et al. (2025) highlighted that exposure to English-language movies significantly contributes to vocabulary acquisition and contextual understanding. Damayanti et al. (2025) pointed out that subtitles serve as an important precursor, helping learners understand words in listening that otherwise might be unheard, functioning as a necessary "external scaffold" for eventual phonological production.

Furthermore, the specific issue of prosodic interference must be addressed to justify the suprasegmental indicators used in this study. Widagsa et al. (2019) provided critical empirical evidence on why pronunciation, specifically Word Stress, remains a major hurdle for Indonesian learners. They argued that the interference of the native Indonesian prosodic system creates a "negative transfer" when learning English, leading to systemic mispronunciation that alters word meanings. Complementing this, Prayitno et al. (2022) highlighted that students frequently struggle with stress placement in polysyllabic words and often rely on digital media to self-correct their errors outside the classroom, a finding echoed by Weda (2018).

Exploring the correction of suprasegmental errors, Toyama and Hori (2025) highlighted the effectiveness of multimodal digital media in reinforcing prosodic structures, including word stress and intonation. They discovered that when students actively engage with visual and auditory stimuli, they are better able to identify auditory and articulatory features, leading to refined pronunciation. Additionally, a study by Jin (2023) investigated the effects of shadowing activities using movies on students' speaking skills, showing significant improvement in intonation and prosody. The shadowing process an imitation task demands high levels of cognitive engagement, which aligns with the

interactive quality of subtitles. This strongly supports the rationale for using a Reading Aloud task as the primary instrument for assessing the dependent variable in this study, as imitation is the most direct test of whether the visual and textual inputs have been internalized.

Despite the extensive literature on multimedia learning, previous studies present notable empirical gaps. Laoli et al. (2025) confirmed the benefits of English movies on speaking skills; however, their study was strictly confined to a literature review methodology, lacking empirical quantitative testing on specific phonetic dimensions. Furthermore, while Baranowska (2020) investigated subtitles, her study was limited to incidental vocabulary learning, explicitly calling for future research on intentional language learning. The present quantitative study fills these precise methodological and empirical gaps by deploying Multiple Linear Regression to examine the intentional phonetic production (pronunciation) of EFL learners exposed to movie clips and subtitles.

I. Synthesis and Research Gap

The synthesis of the aforementioned literature suggests a general consensus that multimedia tools, digital streaming platforms, and subtitles can support English language learning by providing authentic audiovisual input and facilitating learners' engagement. Nevertheless, a methodological and conceptual gap remains evident in the existing body of research.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

Most previous research has either focused on learners' general perceptions through qualitative methods (Putri & Zuraida, 2025; Salsabila, 2022; Damayanti et al., 2025), evaluated broad speaking fluency without phonetic isolation (Laoli et al., 2025), analyzed multimodal phonetic training qualitatively (Toyama & Hori, 2025), or utilized simple experimental designs (pre-test/post-test) without measuring the independent, predictive contribution of specific psychological variables (Mishbahuddin, 2025). Very few studies have operationalized Movie Clip Exposure and Subtitle Awareness as two independent, measurable psychological constructs to examine their natural, correlational contribution to highly specific phonological accuracy.

By employing a Multiple Linear Regression design and a highly objective binary assessment system (Hit or Miss) validated by acoustic speech analysis software (Praat), this study directly addresses these methodological gaps. It does not merely ask "if" movies help; it statistically measures "how much" these two variables individually (partially) and jointly (simultaneously) correlate with seventh-grade students' specific pronunciation skills (Accuracy, Word Stress, and Intonation). This provides precise, discrete quantitative evidence of their predictive relationship, which is largely missing in the current literature concerning Indonesian EFL junior high school contexts.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG