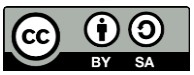
IDEAS

Journal of Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 13, Number 2, December 2025
pp. 9354 - 9367



Issued by English study program of IAIN Palopo

Exploring Students' Perspective of The Use Audio-Visual in English Listening Skill

Nur Rahma Maulida M[1], Munir[2], Geminastiti Sakkir[3], Adeliah Nurmuthmainnah Syam4, Nurfitrah Awaliyah[5]

[1,2,3,4,5]Pendidikan Bahasa Inggris, Pascasarjana Universitas Negeri Makasssar

Corresponding E-Mail: neraklemensia@student.untan.ac.id

Received: 2025-07-28 Accepted: 2025-12-31
DOI: 10.24256/ideas. v13i2.7626

Abstract

This study aims to explore students' perspectives on the use of audio-visual media in learning English listening skills. Employing a descriptive qualitative research design, the study involved six students from Grade XI at SMAN 2 Pangkep. Data were collected through semi-structured interviews and documentation. The findings indicate that students generally hold positive views toward the use of audio-visual media. They perceive it as an effective tool that enhances their engagement, increases motivation, and facilitates a better understanding of listening materials. Students reported that the combination of sound and visual elements helped them grasp meaning more easily, making the learning process more enjoyable and less monotonous. Their perspectives also highlight the role of audio-visual media in supporting active participation, reinforcing comprehension through imagery, and encouraging independent learning. These insights suggest that audio-visual media can be a valuable resource in developing English listening skills from the learners' point of view.

Keywords: *EFL, Audio-visual, Listening skill, Students' perspective*

Introduction

Listening is a fundamental skill in second language acquisition and plays a central role in successful communication. According to Rost (2011), listening is not merely a receptive skill, but an active and complex process involving decoding, interpreting, and constructing meaning from auditory input. In the Indonesian context, English is taught as a foreign language beginning from elementary school, with increasing emphasis on communicative competence at the secondary level (Kebudayaan, 2017). However, the development of listening skills among high school students often remains underdeveloped due to limitations in exposure, instructional methods, and student engagement.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

In recent years, educators have attempted to address these issues by integrating audio-visual media such as films, podcasts, animations, and educational videos into English listening instruction. The rationale for using audio-visual media lies in its multimodal nature, which facilitates better comprehension through visual and auditory reinforcement. Mayer's (2009) *Cognitive Theory of Multimedia Learning* suggests that learners process information more effectively when it is delivered through both visual and verbal channels simultaneously, as this dual coding enhances retention and understanding. In line with this theory, studies such as Seçer et al. (2015) and Sugiani et al. (2020) have shown that the use of AV materials significantly increases learner motivation and enhances language comprehension by providing meaningful and contextualized input.

Despite these promising findings, many EFL students still face considerable challenges when engaging with audio-visual media in listening tasks. Research by Gilakjani and Sabouri (2016) identifies linguistic, cognitive, and technical barriers as common obstacles, such as difficulty processing fast speech, unfamiliar vocabulary, reduced forms, poor sound quality, and a lack of strategic listening awareness. Goh (2000) further argues that EFL learners often struggle with cognitive load during listening due to limited working memory and ineffective decoding strategies. In classrooms where audio-visual media is used, these challenges may persist if students are not guided appropriately in how to navigate and utilize such resources. Thus, while audio-visual tools are believed to enhance learning, their effectiveness can vary greatly depending on how learners interact with them.

More importantly, there is a noticeable gap in the literature regarding students' own perspectives on the use of audio-visual media in developing their listening skills. While previous studies (e.g., Chen, 2020; Wang, 2021; and Lee, 2021) have primarily focused on measuring performance outcomes or comparing traditional and audio-visual-based instruction, fewer studies have investigated how students reflect on their experiences with audio-visual materials in the classroom. Yet, as Dörnyei and Ushioda (2011) emphasize, understanding learners' beliefs, attitudes, and perceptions is crucial because these factors deeply affect engagement, motivation, and learning behaviour. When students are not comfortable with the mode of instruction or perceive it as ineffective, even the most innovative tools may fail to achieve intended learning outcomes.

Additionally, Vandergrift and Goh (2012) highlight the importance of metacognitive listening instruction, which involves raising students' awareness of their listening processes and encouraging reflective learning. Without this dimension, learners may approach audio-visual listening passively—watching and listening without consciously applying strategies to comprehend and retain meaning. Classroom observations in SMAN 2 Pangkep indicate that while students are enthusiastic about using videos or other multimedia in class, they often report difficulties in understanding content due to fast speech, unfamiliar expressions, or lack of clarity in instructions. These preliminary findings suggest that students'

engagement with audio-visual media is not free of problems, and their voices need to be heard in order to understand the pedagogical strengths and weaknesses of this approach.

Thus, this study aims to explore the perspectives of tenth-grade students at SMAN 2 Pangkep regarding their experiences with audio-visual media in English listening lessons. It seeks to identify the benefits, limitations, and obstacles they encounter when using audio-visual materials, with the goal of informing more responsive and student-centred instructional practices. By adopting a qualitative approach, this research contributes to a more nuanced understanding of how audio-visual resources are received by learners in a real classroom context an angle that is often overlooked in experimental or quantitative studies (Graham, 2006). Furthermore, by focusing on a localized Indonesian context, the study offers practical insights into the realities of EFL learning in environments where access to resources and teacher training may be limited.

Ultimately, this research is grounded in the belief that effective listening instruction requires more than just advanced technology or modern media. It requires a deep understanding of how learners interact with, perceive, and are challenged by those tools, and how these tools can be meaningfully integrated into pedagogical practice. By exploring students' perspectives, this study aims to bridge the gap between the theoretical advantages of audio-visual media and the practical experiences of learners, contributing to more adaptive and inclusive teaching strategies in the EFL classroom.

Method

This study employed a descriptive qualitative research design, which was chosen to provide an in-depth and holistic understanding of the reflection encountered by students in using audio-visual materials to develop their English listening skills. The primary aim of this approach is to describe, interpret, and present the phenomena under investigation based on rich, contextual data gathered directly from natural learning environments. As Creswell (2023) explains, qualitative research focuses on exploring and understanding the meanings individuals or groups ascribe to social or human problems. This is done by collecting data in real-life settings, interpreting the findings inductively, and ultimately generating broad themes that reflect participants lived experiences.

Descriptive qualitative research, in particular, seeks to offer straightforward and factual accounts of real-world events. Unlike more interpretative methodologies such as phenomenology, grounded theory, or ethnography, descriptive qualitative research emphasizes accurate depiction of situations, audio-visual, and perspectives without deeply theorizing or abstracting them. In this context, the researcher aimed to describe the learning challenges students face when engaging with audio-visual media, using students' own words and observable audio-visual as primary sources of insight.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

To gather data, the researcher employed two main qualitative techniques: semi-structured interviews and documentation. Both methods were selected to gain a comprehensive understanding of students' perspectives and experiences in using audio-visual materials for developing English listening skills.

Semi-Structured Interviews

Semi-structured interviews were conducted to explore students' personal reflections, challenges, and learning strategies when engaging with audio-visual media in listening activities. This method combines the use of predetermined guiding questions with the flexibility to probe deeper based on participants' responses. It allowed the researcher to maintain focus on key research themes—such as comprehension, motivation, and learning behaviour while also encouraging participants to elaborate on their unique experiences.

The interview protocol consisted of open-ended questions that addressed how students perceived audio-visual materials, what obstacles they faced (e.g., speed of speech, unfamiliar accents, or technical issues), and how they adapted to these challenges. During the interview sessions, students shared how they used visual cues, subtitles, repetition, and context prediction to aid comprehension. Some also reflected on their emotional responses, such as increased interest or frustration depending on the content difficulty. The interviews were recorded and transcribed verbatim to ensure the accuracy of interpretation during data analysis.

Documentation

Documentation was used as a complementary data source to support and verify information obtained through interviews. This method involved the systematic collection and analysis of various learning artifacts and records related to the audio-visual learning process. Key documents included students' worksheets, teacher lesson plans, evaluation rubrics, and reflective notes. These materials provided contextual insight into the instructional design and student responses. Furthermore, the researcher utilized visual documentation, including photographs and video recordings of classroom listening activities involving audio-visual materials.

These were useful in capturing non-verbal behaviours such as attention focus, engagement levels, and collaborative interactions during tasks. By analysing these records, the researcher identified observable patterns such as repeated pausing of video content, peer discussions, or gestures indicating comprehension or confusion. Together, semi-structured interviews and documentation served as mutually reinforcing tools for data triangulation. The comparison between students' verbal reflections and tangible records of classroom practice enhanced the credibility and depth of the findings, allowing for a well-rounded portrayal of how audio-visual media influenced the development of students' English listening skills.

In addition to documentation, semi-structured interviews were conducted to obtain deeper insights into students' personal experiences, perceptions, and difficulties when engaging with audio-visual media for listening activities. This method allowed the researcher to combine a structured format using pre-prepared guiding questions with the flexibility to explore emerging topics based on participants' responses. This approach is particularly suitable for qualitative inquiry, as it encourages participants to elaborate on their thoughts while still keeping the conversation focused on the research objectives.

The interviews were conducted individually with selected participants on the final day of the study. Each session lasted approximately 10 to 15 minutes, providing sufficient time to explore the key themes while maintaining participant comfort and focus. The interviews were audio-recorded (with consent) to ensure accuracy in data transcription and analysis. The spoken responses obtained from the participants provided valuable insights that complemented the observational data and helped triangulate the findings.

In sum, by combining non-intrusive observation with flexible yet guided interviews, this study was able to capture both the observable and experiential dimensions of students' engagement with audio-visual materials. This methodological triangulation strengthened the credibility and richness of the data, enabling a comprehensive exploration of the obstacles faced by learners in developing their listening skills through multimedia-based instruction.

Data Analysis

After collecting data from two instruments, data analysis technique needed to answer the research questions. In analyzing the data, the researcher used interactive model by Miles, Huberman and Saldana (2014).

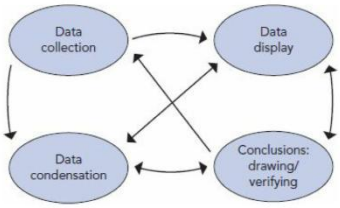


Figure 1. Interactive Model

This model allows researcher to collect the data through writing, editing, classifying, reducing and describing according to this topic. According to Miles, Huberman and Saldana (2014) they presented three streams are data condensation, data display and conclusion drawing/verification.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Data Collection

In collecting data for this study, the researcher used non-participant observation to see the students' perspective learning process using audio-visual. The observation checklist was used as a guide to identify the students' perspective employed by the students during learning process. The end of the study, the researcher used the semi-structured interviews conducted after the learning process is completed. To support the data accuracy, the researcher used documentation in the form of photos and videos during the teaching process, as well as audio recordings during interviews with the teachers.

Data Condensation

Data condensation means to summarize, sorting out the main things and looking for the themes/patterns. Miles, Huberman, and Saldana (2014) assert that data condensation involves the selection, focus, simplification, abstraction, and/or transformation of data from the comprehensive collection of written field notes, interview transcripts, documents, and other empirical materials. Data condensation involves composing summaries, coding, formulating themes, creating categories, and drafting analytic notes. In this study, data condensation serves to filter and simplify the raw data from observations and interviews, concentrating on three main aspects: opening, main activity, and closing. After conducting in-depth interviews, irrelevant segments (such as casual conversation) are excluded, keeping only the relevant responses for further analysis.

Data Display

Data display refers to a visual or narrative presentation of the processed data, where relevant and important information is selected and organized in a way that facilitates further analysis. The data analysis follows the design and procedure of the study. This stage highlights key findings by arranging relevant data in a structured format such as charts, matrices, diagrams, or descriptive narratives that helps researchers identify patterns, draw conclusions, and answer research questions effectively. A clear and well-organized data display allows for deeper insight into the phenomenon being studied and supports the overall validity of the analysis.

Drawing and Verifying Conclusions

The last step is drawing conclusion and verification from the information obtained from the field. The researcher made conclusions based on the completed analysis and double-checked the selected data by comparing it with the evidence collected in the field.

Results

Based on the results of the observations and interviews conducted by the researcher, it was found that three stage of factors that faced by students' perspective aligns with Bruner's (1996) theory when learning English using Audio-visual are Enactive Representation, Iconic Representation, and Symbolic Representation.

Enactive Representation

The findings of this study indicate that the enactive representation stage significantly enhances students' listening comprehension when learning English through audio-visual materials. Students demonstrated improved understanding and vocabulary retention by physically engaging with the instructions presented in the videos. For example, one participant stated, *"Senangka bejalar seperti ini karena sambil nonton video, guru juga kasih instruksi dari video, terus sudah itu bergerakki sesuai sama petunjuknya. Kayakka kurasa main petualangan"* ["I like this kind of learning because while watching a video, the teacher also gives instructions from the video, then we move according to the instructions. It feels like playing an adventure."].

Another student explained, *"lebih mengerti ma perbedaan 'go straight' dan 'turn left' karena saya langsung praktek jalan waktu dikasih perintah. Jadi ceritanya bukan cuman menonton, tapi bergerakki juga"* ["I understand the difference between 'go straight' and 'turn left' better because I immediately practiced walking when given the command. So, the story is not just watching, but moving too."]. These embodied learning experiences not only fostered active engagement but also enabled students to internalize directional vocabulary more effectively. The combination of seeing, hearing, and physically responding to the input created a context-rich environment where comprehension was reinforced through action, aligning well with Bruner's concept of enactive representation

Iconic Representation

The results of this study demonstrate that the use of iconic representation in audio-visual materials significantly supports students' comprehension of directional vocabulary in English. Visual cues such as arrows, animated characters, and icons of places like shops or schools serve as concrete references that help students interpret abstract linguistic input. For instance, one student shared, *"Pasku lihat video, seketika kutaumi artinya 'next to' karena ada gambar toko di sebelah kantor pos. Terus bisama pahami kata itu cuman lihat saja gambarnya"* ["When I saw the video, I immediately understood what 'next to' meant because there was a picture of a shop next to the post office. Then I could understand the word just by looking at the picture."].

Another participant remarked, *"Pas lihat animasi jalur jalan, langsungka tahu bedanya 'turn around' dan 'go back'. Karena da gambaranya toh kak, kuliatki mobilnya putar balik, jadi saya bisa langsukung ku bedakan dua vocab itu"* ["When I saw the road animation, I immediately knew the difference between 'turn around' and 'go back'. Because there was a picture of the car making a U-turn, so I could immediately distinguish between the two vocab."].

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

These extracts highlight how the visual context enables learners to decode and internalize meaning more effectively without needing further verbal explanation. Additionally, the presence of consistent visual symbols, such as directional arrows or distinct building icons, facilitates rapid recognition and retention of vocabulary. As another student explained, *"Ku perhatikan ki di video pasti di setiap tempat seperti supermarket atau police station punya ikon berbeda. Itu bikin saya cepat mengenali kata-kata itu karena otak saya langsung hubungkan dengan gambarnya"* ["I noticed in the video that every place like a supermarket or police station has a different icon.

That makes me quickly recognize the words because my brain immediately connects them with the image."]. Overall, these findings affirm that iconic representation enhances listening comprehension by linking spoken input with clear, contextually relevant visual images, enabling students to form mental associations that support long-term understanding.

Symbolic Representation

Apart from enactive representation and iconic representation, another indicate that symbolic representation plays a crucial role in helping students advance their English listening skills through audio-visual materials, particularly in the context of giving directions. At this stage, students begin to internalize language structures and vocabulary to the extent that they can comprehend and produce responses without relying on visual cues. For example, one student noted, *"Sedikit percaya dirija kak, karena bisama pahami arti kalimat seperti 'Go straight and turn left at the bank' tanpa lihat video lagi, karena saya sudah hafal kosakatanya itu"* ["I can already understand the meaning of sentences like 'Go straight and turn left at the bank' without looking at the video again, because I have memorized the vocabulary."], reflecting increased confidence and autonomy.

Another student added, *"Kalau gurunya kasih kalimat arah secara lisan, saya bisa langsung tulis ulang dalam bentuk kalimat tertulis"* ["If the teacher gives me directions verbally, I can immediately rewrite them in written form."], which demonstrates the ability to convert spoken input into written symbolic output. This linguistic competence was further reinforced through tasks such as dialogue creation and command comprehension, with students articulating how they learned to use key vocabulary like *'between', 'across from',* and *'turn around'* accurately. Additional insights reveal that students can now complete listening tasks without visual aids, as expressed in *"Bisama mengerjakan soal listening tentang direction tanpa lihat gambar kak, karena dari video itu, sudah ku tahu mi arti dan pola kalimatnya"* ["I can do listening problems about direction without looking at pictures, because from the video, I already know the meaning and sentence patterns."].

Collectively, these extracts confirm that symbolic representation has enabled learners to process verbal directions meaningfully, reflect them in both spoken and written communication, and apply them independently—showing that the use of audio-visual materials not only supports earlier learning stages (enactive and iconic) but also builds toward higher-level, language-based thinking.

Discussion

The primary objective of this study was to explore and analyse students' perspectives in developing their English listening skills through the integration of audio-visual materials in classroom settings. This qualitative investigation, conducted through in-depth interviews and direct classroom observations, revealed critical insights into how learners experience and internalize English language input when exposed to multimodal resources. Drawing on Jerome Bruner's theoretical framework of cognitive representation—namely, enactive, iconic, and symbolic stages—this study categorized student responses and behaviours according to the cognitive processes they exhibited during audio-visual learning activities.

Findings indicated that students navigated these three representational stages progressively, starting from physical engagement with language, advancing through visual comprehension, and ultimately reaching abstract symbolic mastery. Each stage contributed uniquely to learners' acquisition and application of listening skills, vocabulary, and sentence construction. The following discussion elaborates on each stage of Bruner's model and connects it with the observed learning outcomes and experiences of the students.

a. Enactive Representation

Enactive representation, which constitutes the first and most tangible level in Bruner's model of cognitive development, involves the acquisition of knowledge through direct physical action. In this stage, learners associate meaning with bodily movements and actions rather than with symbols or images. The findings from this study strongly reflect the effectiveness of enactive representation in supporting students' early language comprehension, particularly in listening activities related to giving and following directions.

During classroom sessions, students were often asked to physically follow oral commands such as "turn left," "go straight," or "walk past the bank." These directions were accompanied by animated videos or situational simulations that prompted students to act out the instructions in real time. One student specifically noted, "Langsung mi kuhafal karena sambil praktek," highlighting the immediacy of vocabulary retention when action was paired with instruction. Such kinesthetic engagement is consistent with cognitive learning theories that emphasize the importance of sensorimotor integration in strengthening memory and understanding (Anderson, 2005).

Moreover, these physical responses were not always initiated by teachers but were often self-directed by students. This autonomy suggests an emerging metacognitive awareness, as learners began to identify strategies that helped them process language input more effectively. Several students described the learning process as enjoyable and likened the activities to "playing an adventure." These emotional reactions—stemming from active involvement and playful interaction—

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

are important, as positive emotions are known to improve memory consolidation and increase motivation to learn (Vygotsky, 1978)

The physical embodiment of listening tasks also helped reduce cognitive load. Rather than relying solely on auditory decoding, students were able to process meaning through movement, which served as an additional support mechanism. This multimodal reinforcement allowed for a deeper cognitive connection between the language being heard and the physical action being performed, making learning both efficient and memorable. As a result, enactive representation not only facilitated immediate comprehension but also laid the foundation for more complex understanding at the visual and symbolic levels.

In short, enactive learning empowered students to engage with English listening material in a meaningful and memorable way. It promoted learner autonomy, enhanced motivation, and created an emotionally positive and cognitively supportive environment that made language acquisition less intimidating and more experiential

b. Iconic Representation

The second phase of Bruner's model, iconic representation, involves learning through the use of images and visual associations. In this stage, learners begin to link language with visual stimuli, thereby transitioning from concrete physical experiences to visual-symbolic comprehension. In the context of this study, iconic representation played a central role in helping students make sense of spoken English through animated visuals, symbols, and pictorial prompts.

Students were exposed to various video materials that visually illustrated everyday contexts—maps, shops, road signs, and directional arrows. When listening to instructions such as "next to the school" or "behind the supermarket," students reported a clearer understanding once these phrases were paired with corresponding images. For instance, in Extract 29, a student stated that upon seeing a shop placed next to a post office on a video map, she could immediately grasp the meaning of the phrase "next to." Similarly, terms like "turn around" or "go back" were understood more easily when supported by animations showing characters moving accordingly.

This aligns with Mayer's (2001) Cognitive Theory of Multimedia Learning, which posits that people learn better from words and pictures than from words alone. Students processed information through both visual and auditory channels, reducing extraneous cognitive load and enabling more effective comprehension. The dual-channel processing allowed learners to visualize what they were hearing, which contributed to deeper understanding and more efficient retention.

Another significant advantage of this visual modality was the use of consistent symbolic icons—such as hospitals, schools, and supermarkets—that repeatedly appeared across different learning scenarios. These recurring images helped reinforce vocabulary recognition. Students began to associate specific visual symbols

with their English counterparts, enabling long-term recall even when visuals were no longer present. This form of image-based repetition proved to be a scaffold that supported students' progression toward symbolic representation, as it helped solidify word meanings through visual memory.

Furthermore, iconic representation offered a bridge for students who struggled with purely auditory input. Those who initially had difficulty understanding spoken directions found that pairing the language with visuals provided necessary support, thereby increasing their confidence and willingness to engage. The visual aids thus acted not only as cognitive tools but also as emotional buffers, helping to build learner confidence and reduce anxiety commonly associated with foreign language listening tasks.

Therefore, iconic representation functioned as a crucial transitional phase in students' learning journeys. It helped translate auditory information into recognizable visual patterns, allowing learners to make meaning through association and pattern recognition. As students became more adept at decoding both sound and image, their capacity for abstract language processing grew, setting the stage for the final phase: symbolic representation.

c. Symbolic Representation

The final stage in Bruner's model—symbolic representation—marks the highest level of abstraction, where learning takes place through the manipulation of symbols such as language, numbers, or logical constructs. In this study, symbolic representation was evident in students who were able to comprehend complex instructions in English without relying on physical actions or visual support.

Students demonstrated symbolic competence when they responded accurately to purely verbal commands, such as "Turn left at the traffic light, then go straight for two blocks and stop near the post office." Even without visual prompts, they could follow the instructions and describe the movement sequences clearly. This suggests that learners had internalized not just vocabulary but also the grammatical structures and spatial relationships embedded within the language. It is this internalization that signifies a shift from experiential understanding to conceptual mastery.

Moreover, symbolic proficiency was also observed when students were asked to respond to audio-only prompts or to write dialogues and short paragraphs based on listening tasks. Their ability to produce coherent responses without relying on images or actions indicates that they were no longer dependent on external cues. Instead, they had developed an internal mental model of the language, which enabled them to interpret, manipulate, and generate new content in English.

This stage also revealed students' growth in metacognitive and linguistic skills. They could reflect on their own learning process, recognize patterns in language input, and apply previously learned structures in novel contexts. This kind of abstract thinking is the hallmark of symbolic representation, as it requires

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

learners to make connections between different cognitive domains—such as listening, reading, and writing—without the aid of concrete referents.

Importantly, students themselves attributed this ability to their repeated exposure to audio-visual materials during earlier stages. The foundational work done through enactive and iconic representations provided a scaffold upon which symbolic competence could be built. Students who had previously enacted directional commands or visualized location-based instructions were now capable of processing those same concepts purely through auditory or textual input. This developmental trajectory affirms Bruner's theory that learning should progress from concrete to abstract, with each stage building upon the previous one.

Symbolic representation, therefore, represents not just a milestone in language learning but a transformation in how students relate to language itself. At this stage, language is no longer something to be acted out or visualized—it becomes a tool for thought, communication, and problem-solving. It empowers learners to operate independently, think critically, and engage with English in both academic and real-world contexts.

Conclusion

This study aimed to explore students' perspectives on the use of audio-visual media in enhancing English listening skills, using Bruner's theory of cognitive representation as a conceptual framework. Through a descriptive qualitative approach, the research identified how learners internalized listening input across three progressive stages—enactive, iconic, and symbolic representation. Each stage was evidenced through classroom observations and students' interview responses, demonstrating a clear transformation from concrete to abstract understanding in their engagement with English directional language.

At the enactive level, students benefitted from physical engagement acting out directions, mimicking movements, and responding kinaesthetically to commands presented through video content. These bodily experiences helped anchor new vocabulary and sentence structures in real-world actions, increasing retention and comprehension. This physical interaction also contributed to a sense of enjoyment and reduced anxiety, making the learning process more accessible and less intimidating for beginners.

Next there is the iconic stage showed how visual input played a pivotal role in connecting auditory language with meaningful images. Learners relied on video animations, maps, and contextual cues to form mental associations between English expressions and their corresponding real-life objects or scenarios. These visuals allowed students to decode unfamiliar vocabulary and sentence patterns, eventually leading to more fluent understanding of spoken language. Furthermore, visual media supported students who may struggle with purely auditory learning by offering them a second, reinforcing channel of input.

Last one is symbolic representation, as the highest cognitive stage, revealed students' growing independence in understanding and using English without reliance on visual or physical supports. Here, learners demonstrated the ability to interpret and reconstruct meaning from verbal language alone, whether by writing responses, constructing dialogues, or recalling directions solely based on listening. Their ability to move across modalities—from listening to writing, or from audio input to verbal output—reflects the development of abstract thinking and deeper linguistic mastery.

The integration of audio-visual materials, therefore, is not merely a tool for delivering content but a catalyst for cognitive and linguistic transformation. By activating multiple representational modes, these materials scaffold learners from surface-level exposure to meaningful language use. The findings suggest that well-designed audio-visual content can foster not only comprehension but also motivation, confidence, and communicative competence in English as a foreign language.

In sum, audio-visual media support a comprehensive learning journey, enabling students to construct knowledge through action, association, and abstraction. Educators should recognize the pedagogical value of these materials and consider embedding them systematically into listening instruction. A thoughtful combination of video-based tasks, interactive physical responses, and follow-up symbolic exercises—such as dialogue writing or retelling—can promote richer language acquisition and bridge the gap between understanding and application.

Future research might explore how these representational stages unfold over time and how they differ across various learner profiles, language levels, or types of listening content. Ultimately, the study confirms that listening instruction enriched by audio-visual input can significantly enhance students' ability to engage with English meaningfully, purposefully, and independently

References

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). Worth Publishers.

Bruner, J. S. (1996). *The culture of education*. Harvard University Press.

Chandra, J., & Linuwih, E. R. (2024). Female Language Features in Women's Twitch App Livestreaming. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(2), 1554-1566.

Gilakjani, A. P., & Sabouri, N. B. (2016). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching, 9*(6), 123–133. https://doi.org/10.5539/elt.v9n6p123

Lestari, N. P. D. E., Artini, L. P., & Adnyani, N. L. P. S. (2025). An Analysis of Code Mixing and Code Switching Used in Maudy Ayunda's YouTube Channel Video. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(1), 2292-2306.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164603

Miles, B. Mathew dan Michael Huberman. 1992. Analisis Data Kualitatif Buku Sumber Tentang Metode-metode Baru. Jakarta: UIP.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Muhammad, B. S., & Setiawan, R. (2025). Conventional Pedagogies in EFL Higher Education: A Critical Inquiry into Lecturers' Beliefs and Student Voices. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(1), 3062-3072.

Novarita, P., satya Gumelar, W., & Ainun, N. (2025). The Influence of English Youtube Videos on The Disoriented Language of The Mother Tongue Among Young Children on Their Language Development Process. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(1), 2995-3013.

puspita Candra, K. D. (2024). An Analysis of Directive Speech Acts Found in "The Expendable 2" Movie. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(2), 2404-2414.

Pranata, I. P. Y., & Putri, I. G. A. V. W. (2024). Expressive Illocutionary Acts found in Spider-Man: Across the Spider-Verse Movie. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(2), 2298-2313.

Rachmawati, A. (2024). YouTube Video in Improving Senior High Students' Vocabulary Mastery. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(2), 2204-2220.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.