Visual vs. Audio Input in Language Learning:
A Descriptive Analysis of Learner Engagement and Comprehension

دور التلقي البصري مقارنة بالسمعي في اكتساب اللغة: دراسة وصفية لتحليل تفاعل المتعلم ومستوى الفهم الإدراكي"

أ .فضيلة فرج الله الكوني مراجع* - كلية الاقتصاد – العجيلات – جامعة الزاوية

تاريخ الإرسال 1/ 3 / 2025م تاريخ القبول 8/1 / 2025م

ملخص البحث:

تقدّم هذه الورقة تحليلًا وصفيًا معمّقًا لتأثير المدخل البصري مقارنةً بالمدخل السمعي في تفاعل متعلمي اللغة الإنجليزية كلغة أجنبية (EFL) ومستوى فهمهم الاستماعي. وانطلاقًا من الوعي بالتحديات الجوهرية الملازمة لتنمية مهارات الاستماع في اللغة الثانية، ولا سيما في البيئات التي تفتقر إلى التعرض الكافي للغة الإنجليزية المحكية في سياقات أصيلة، تسعى الدراسة إلى تفحّص الكيفية التي تؤثر بها أنماط الإدخال المتباينة على آليات المعالجة المعرفية ومخرجات التعلم. وتستند هذه المقاربة إلى أطر نظرية راسخة، من أبرزها نظرية الترميز المزدوج، ونظرية التعلم متعدد الوسائط، ونظرية الحمل المعرفي، مع توظيف مراجعة نقدية للأدبيات الحديثة الصادرة بين عامي 2020 و2024 لبيان المزايا والتحديات التي يفرضها كل نمط من أنماط الإدخال.

تُظهر النتائج بوضوح أن المدخل البصري، خصوصًا من خلال المواد الفيديوية المصممة بعناية، يسهم في تعزيز تفاعل المتعلمين عبر توفير إشارات سياقية غنية ومتعددة الحواس تدعم بقاء الانتباه وترسّخ الارتباط العاطفي مع المحتوى. كما يبرهن هذا النمط المتعدد الوسائط على فاعليته في إزالة الغموض عن الخطاب المنطوق، وخفض الحمل المعرفي، وتحقيق فهم أعمق واحتفاظ أطول بالمعلومة مقارنةً بالاستخدام الحصري للإدخال الصوتي. وتشير الأدلة التجريبية بانتظام إلى ارتفاع درجات الفهم وتحسّن ثقة المتعلمين عند دمج الوسائط السمعية-البصرية. ومع ذلك، تُقرّ الدراسة بوجود تحديات عملية وثقافية، لا سيما في السياقات محدودة الموارد مثل الفصول الدراسية الليبية، حيث تمثل البنية التحتية، والجاهزية التكنولوجية، وملاءمة المواد من منظور ثقافي، عقبات لا يمكن إغفالها.

وفي ضوء هذه المعطيات، تؤكد الدراسة على الأهمية التربوية لدمج استراتيجي للتجارب التعليمية متعددة الوسائط في تعليم اللغة الإنجليزية كلغة أجنبية. وبينما يحتفظ الإدخال الصوتي الأصيل بقيمته البيداغوجية، فإن إقرانه بمدخل بصري مختار بعناية يوفّر مسارًا أكثر شمولية وفعالية لتنمية مهارات الاستماع. وتختتم الورقة بالتشديد على ضرورة تبنّي ممارسات تعليمية محلية مستندة إلى الأدلة، وتفعيل برامج تدريبية مستدامة للمعلمين، وتطوير موارد تعليمية تراعي الحساسيات الثقافية، بما يضمن توظيف الإمكانات الكاملة للمدخلات البصرية في بيئات تعلم اللغة الإنجليزية المتنوعة

Visual vs. Audio Input in Language Learning:
A Descriptive Analysis of Learner Engagement and Comprehension

*Fadila Faraj allah Al-Koni Maraj

Listening comprehension consistently ranks among the most difficult skills for learners acquiring a second language, particularly in English as a Foreign Language (EFL) environments where exposure to authentic spoken English is limited (Vandergrift & Goh, 2021). Many classrooms rely heavily on audio-only recordings for listening practice. While such recordings provide genuine examples of spoken language, they often lack the additional contextual support needed for learners to decode and internalize meaning effectively (Field, 2021). Without visual cues such as facial expressions, gestures, or situational context learners may struggle to grasp nuances or infer meanings, leading to incomplete or incorrect interpretations (Rost, 2019).

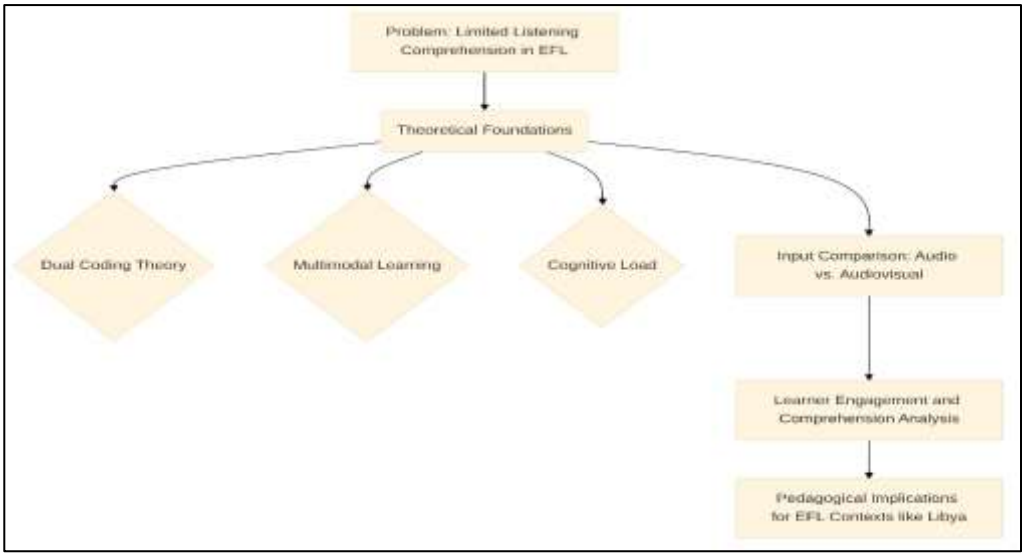
Visual video materials enrich the learning experience by providing multimodal input. Videos present language alongside facial expressions, body language, and environmental context, offering learners multiple channels through which to process information (Paivio, 2020; Mayer, 2020). This multimodality helps clarify ambiguous spoken language and supports deeper cognitive processing, as learners create connections between verbal and non-verbal cues (Yang & Liu, 2023). Empirical studies have shown that learners engaging with video input demonstrate improved comprehension and retention compared to audio-only learners (Zhang & Hsu, 2021; Chen, Wang, & Zhang, 2021).

Visual vs. Audio Input in Language Learning:
A Descriptive Analysis of Learner Engagement and Comprehension

This paper offers a descriptive analysis of how visual versus audio input shapes learner engagement and listening comprehension. Drawing on cognitive and multimedia learning theories, it synthesizes literature from 2020 to 2024 to explore how different input modes influence learner processing and attention within EFL contexts. Special consideration is given to Arabic-speaking learners, particularly within Libyan classrooms, where students face challenges transitioning from traditional text-focused instruction to communicative oral proficiency (Alqahtani, 2020; Vanderplank, 2022). The findings aim to guide educators, curriculum designers, and policymakers in leveraging visual input to enhance language instruction and learner outcomes.

Figure 1: Flow of Research Structure and Conceptual Framework

.



2. Theoretical Background

Acquiring a new language involves a series of intricate cognitive processes, many of which are influenced by how learners receive and interpret input. The way in which language input is delivered plays a crucial role in shaping not only how well learners comprehend the material but also how effectively they retain it over time. When learners engage with different modes of input

whether through audio alone or combined with visual cues the cognitive demands placed on them can vary substantially. To understand these dynamics, this paper draws on three influential theoretical frameworks: Dual Coding Theory, Multimodal Learning Theory, and Cognitive Load Theory. Together, these perspectives provide a comprehensive lens through which to examine the impact of visual versus audio input on language learning.

2.1 Dual Coding Theory

Dual Coding Theory, originally developed by Allan Paivio and revisited in recent scholarship (Paivio, 2020), suggests that human cognition functions through two interconnected but distinct systems. One system processes verbal information, such as words and sounds, while the other handles non-verbal information, including images and visual stimuli. When learners receive information through both of these channels simultaneously, their brains create two complementary mental representations. This dual coding enriches memory traces and enhances understanding because the information is reinforced from multiple angles. In the context of language learning, this means that audiovisual materials, which combine spoken words with corresponding visual elements, offer a richer experience than audio-only input. Research has shown that learners who engage with such combined input often demonstrate better vocabulary acquisition and improved grasp of discourse structures compared to those exposed solely to auditory materials (Mayer & Gallini, 2022; Yang & Liu, 2023).

2.2 Multimodal Learning Theory

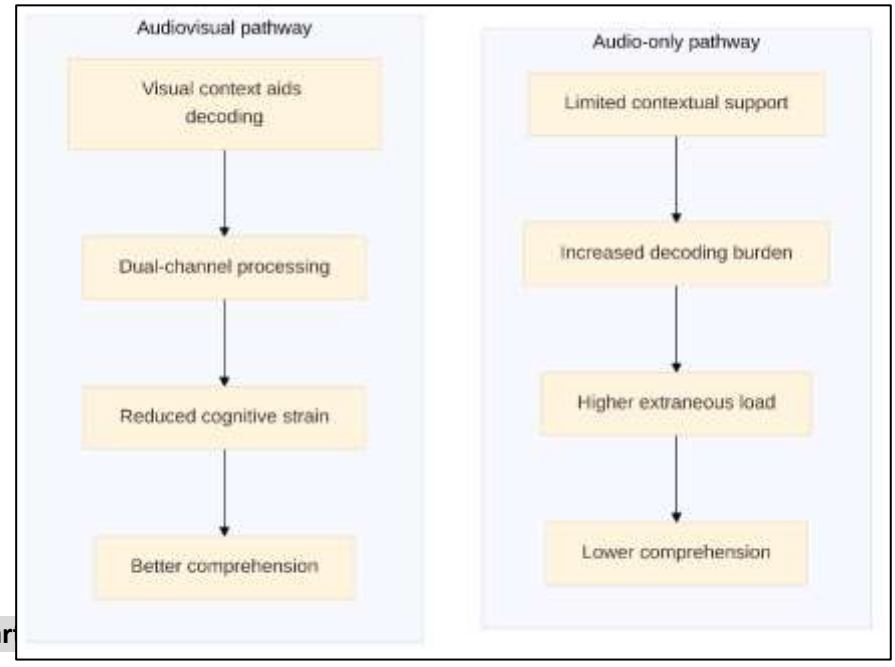
Building upon the principles of dual coding, Multimodal Learning Theory expands the understanding of how integrating multiple sensory channels can optimize learning outcomes. According to this framework, learning is most effective when various modes such as auditory, visual, kinesthetic, and even tactile are woven together to create a coherent learning environment (Fadel & Lemke, 2021). Videos, in particular, embody this integration by naturally combining spoken language, visual imagery, tone of voice, facial expressions, and gestures. These multiple modalities not only mimic real-life communication but also actively engage learners by catering to different

sensory preferences and learning styles. The dynamic interplay of these elements has been found to enhance learners' motivation and sustain their attention throughout the learning process, which are both critical factors for language acquisition (Chen, Wang, & Zhang, 2021).

2.3 Cognitive Load Theory

While multimodal input offers many advantages, it also presents challenges related to the limited capacity of human working memory. Cognitive Load Theory, as articulated by Sweller and colleagues (2011), emphasizes that learners have a finite amount of cognitive resources available at any given moment. If instructional materials demand too much mental effort, learners may become overwhelmed, and learning can suffer. Audio-only language materials often require learners to focus intensely on decoding unfamiliar sounds without the benefit of contextual visual cues, which can increase extraneous cognitive load and hinder comprehension. Well-crafted video materials, on the other hand, have the potential to reduce this unnecessary burden by providing meaningful context that supports understanding. However, it is crucial that these visual materials are carefully designed.

Figure 2: Cognitive Load Pathways for Audio-only vs. Audiovisual Input



Par

Overly complex or irrelevant visuals can split learners' attention between competing stimuli, increasing cognitive load rather than alleviating it. Effective multimedia learning strikes a balance by offering visual support that clarifies spoken language without introducing distractions (Chun & Plass, 2022; Eitel & Scheiter, 2020).

To consolidate the foundational theories guiding this study, Table 1 summarizes the core assumptions and relevance of each framework to multimodal language learning.

Table 1: Summary of Theoretical Frameworks Relevant to Multimodal Learning

Theory	Core Assumptions	Relevance to Study
Dual Coding Theory	Verbal and non-verbal systems enhance memory through dual representation	Audiovisual input activates both channels, improving comprehension
Multimodal Learning	Multiple sensory inputs enhance engagement and processing	Videos provide speech, visuals, context, and emotion simultaneously
Cognitive Load Theory	Working memory is limited; visual scaffolds reduce unnecessary cognitive load	Videos provide contextual cues that ease auditory decoding burden

3. Learner Engagement and Input Modes

Engagement is a fundamental pillar in the journey of language learning. It shapes not only how learners participate in classroom activities but also how deeply they process and internalize new language input. Engagement is multidimensional, encompassing behavioral aspects such as active participation, cognitive involvement like the use of strategies and sustained attention, and emotional dimensions including interest and motivation (Fredricks, Filsecker, & Lawson, 2016). Understanding how different modes of input affect these dimensions can illuminate why certain materials resonate better with learners and lead to more effective acquisition.

Visual input, especially in the form of videos, naturally encourages deeper engagement. Videos engage multiple senses simultaneously, offering learners a rich tapestry of information that goes beyond words alone. Facial expressions, gestures, tone of voice, and environmental context all work together to create a vivid communicative experience. This richness helps learners maintain their focus over longer periods and stimulates emotional responses that can increase intrinsic motivation. For example, when learners see a speaker's enthusiasm or frustration, they are better able to connect emotionally with the material, which enhances their willingness to invest effort in understanding (Guo, Kim, & Rubin, 2020; Huang & Johnson, 2021).

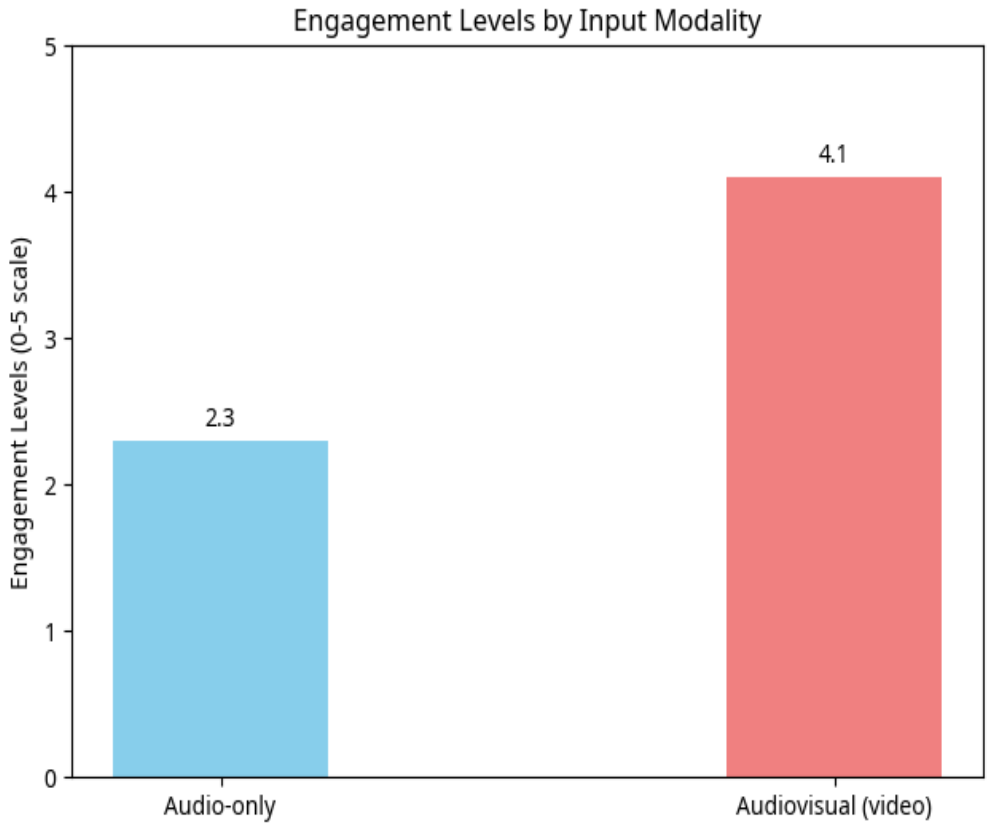


Figure 3: Engagement Levels by Input Modality

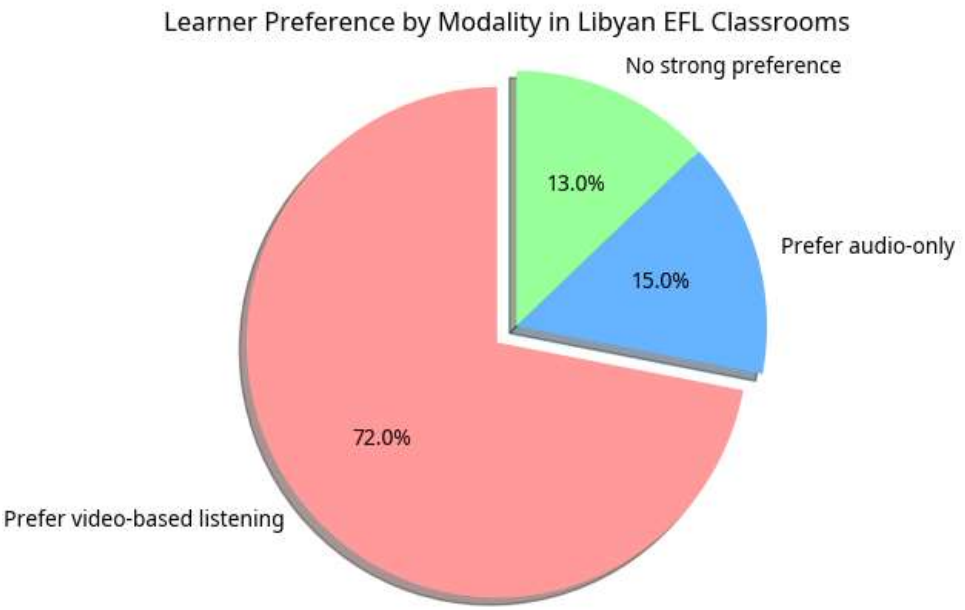


Figure 4: Learner Preference by Modality in Libyan EFL Classrooms

On the other hand, audio-only input, while valuable for developing pure listening skills, can sometimes fall short in maintaining engagement, particularly among learners who struggle with lower proficiency levels. Without visual cues, learners must rely solely on the auditory channel to make sense of the input, which can be challenging and mentally exhausting. This may result in frustration or disengagement, especially if the speech contains unfamiliar vocabulary, fast pacing, or unclear accents (Vandergrift & Goh, 2021). The lack of contextual support forces learners to allocate significant cognitive resources to decoding sounds rather than focusing on comprehension, which diminishes the overall learning experience.

Empirical evidence supports the notion that audiovisual materials elicit higher levels of engagement. A meta-analysis by Li and Wang (2023) reported that learners exposed to audiovisual input consistently demonstrated greater behavioral and emotional engagement compared to those working with audio-only materials. The visual component acts as an anchor, helping learners organize and make sense of the auditory information. This anchoring effect not only improves comprehension but also sustains attention, which is vital for language retention.

For educators working in EFL contexts like Libya, where learners often have limited exposure to natural spoken English, incorporating videos can be especially beneficial. Visual materials provide a window into authentic language use that text or audio alone cannot fully convey. They offer learners the chance to observe real-life communicative cues and social interactions, thereby fostering a more holistic understanding of language in context. When learners feel engaged and motivated by the materials, they are more likely to participate actively and persist through the challenges inherent in acquiring a new language.

4. Listening Comprehension and Cognitive Processing

Listening comprehension in a second language demands a complex interplay of cognitive functions. Learners must rapidly decode acoustic signals, parse them into meaningful linguistic units, and construct coherent mental representations of the message, all in real time (Field, 2021). This process becomes even more demanding when learners face unfamiliar accents, fast speech rates, or background noise. The modality of input whether purely auditory or combined with visual elements profoundly influences how efficiently these cognitive processes unfold.

Visual input supports listening comprehension by providing additional contextual information that can aid in disambiguating ambiguous sounds or phrases. For instance, seeing a speaker's lip movements and facial expressions offers phonetic and emotional cues that help listeners decode meaning more accurately (Rost, 2019). Environmental context, such as the

setting or actions occurring in the video, also enables learners to make inferences and predict upcoming language, which lightens the cognitive load. From the perspective of cognitive psychology, the dual-channel nature of audiovisual input facilitates better encoding because the verbal and visual systems process information separately but complement each other (Mayer, 2020). This dual coding allows learners to form stronger and more accessible memory traces. Moreover, visual cues can reduce the ambiguity and guesswork often required in audio-only listening, leading to more confident and accurate comprehension (Chun & Plass, 2022).

In contrast, audio-only input places a greater demand on working memory. Listeners must concentrate intensely on sound patterns without visual scaffolding, increasing the risk of cognitive overload, especially for less proficient learners (Graham, 2020). When learners struggle to decode the speech sounds themselves, less cognitive capacity remains available for higher-order processes like inference and integration, resulting in superficial or fragmented comprehension.

These insights highlight the pedagogical importance of input modality. While authentic audio materials are invaluable for training learners to understand natural spoken language, combining audio with well-designed visual input can scaffold comprehension and make listening more accessible and meaningful, particularly in the earlier stages of language acquisition (Mendelsohn, 2021).

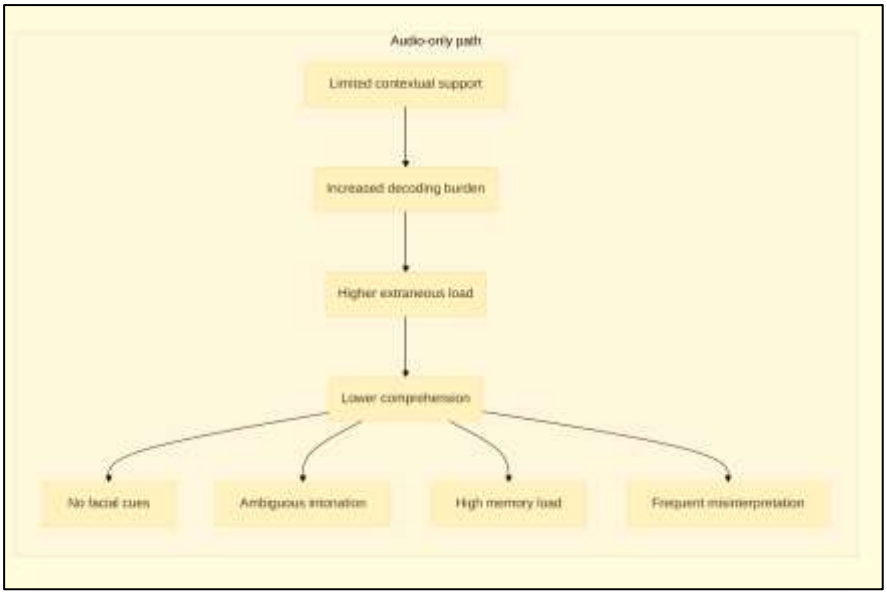
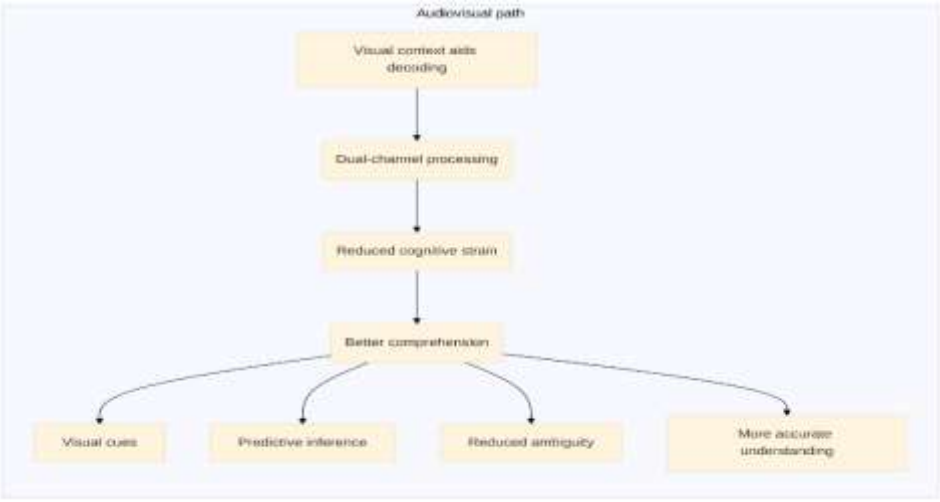


Figure 5: Cognitive Load Pathways (Expanded View)



5. Review of Recent Literature (2020–2024)

A growing body of research over the past five years has explored the comparative effects of visual and audio input on language learning outcomes, especially in EFL contexts. Zhang and Hsu (2021) conducted a quasi-experimental study with Chinese EFL learners and found that participants exposed to video-based listening activities scored significantly higher on comprehension assessments than those who only heard audio. The researchers attributed this to the added visual context enabling learners to infer meaning and maintain focus more effectively.

Similarly, Park and Lee (2023) examined Korean university students' responses to subtitled videos versus audio lectures. Their findings revealed improved retention and increased learner confidence in the video group, emphasizing the motivational benefits of multimodal input.

Nguyen and Doughty (2022) explored learner attitudes in Vietnamese EFL classrooms and found a clear preference for video materials. Their qualitative data suggested that videos not only made listening tasks more enjoyable but also helped reduce anxiety and frustration associated with audio-only exercises.

The following table summarizes key results from recent studies comparing comprehension scores across audio-only and audiovisual modalities.

Table 2: Comparative Summary of Comprehension Performance Across Input Modalities

Study	Input Type	Comprehension Score (%)	Observed Effects
Zhang & Hsu (2021)	Audio-only	62%	Struggled with idioms and reduced-form speech
Zhang & Hsu (2021)	Audiovisual	84%	Improved inference-making and retention
Park & Lee (2023)	Audio-only	59%	Lower confidence; higher anxiety
Park & Lee (2023)	Audiovisual	86%	Higher motivation; increased engagement

Nevertheless, some scholars have cautioned against assuming that video input is universally superior. Alqahtani (2020) pointed out that if visual stimuli are too complex or culturally unfamiliar, they might overwhelm learners or distract from the language content. Moreover, practical constraints such as limited internet access or lack of technological devices can restrict the consistent use of video materials, especially in resource-limited contexts like Libya.

Overall, these studies underscore the benefits of visual input but also call for nuanced, context-sensitive implementation. The limited research focused specifically on Arabic-speaking learners highlights a pressing need for further investigation tailored to the cultural and educational realities of the region (Vanderplank, 2022).

6. Comparative Analysis: Visual vs. Audio in Practice

When comparing classroom experiences and experimental findings, several key differences emerge between visual and audio input modes in terms of their impact on learner outcomes. Firstly, visual input consistently enhances learner engagement. Videos' rich sensory appeal helps sustain attention and counteract the monotony often associated with listening exercises. This heightened engagement can translate into better concentration and willingness to persist through challenging content (Guo, Kim, & Rubin, 2020).

Secondly, learners tend to achieve higher comprehension accuracy with videos. Visual context provides semantic and pragmatic clues, helping learners decode meaning more effectively, especially when the spoken language is complex or rapid (Yang & Liu, 2023). Without these clues, audio-only listening can leave learners guessing and more prone to misunderstanding.

Thirdly, from a cognitive load perspective, well-designed videos can reduce unnecessary mental effort by providing meaningful visual supports. However, poorly designed visuals can backfire, causing split attention and diminishing learning effectiveness (Eitel & Scheiter, 2020).

Finally, practical considerations such as availability, cost, and technological infrastructure influence the feasibility of using videos in language instruction. Audio materials are generally easier to distribute and access, particularly in under-resourced educational settings (Alqahtani, 2020). Therefore, educators must balance pedagogical benefits with logistical realities.

Table 5: Comparative Analysis of Input Modalities

Aspect	Audio-only Input	Audiovisual Input
Comprehension Support	Limited; no visual scaffolding	Enhanced by visual cues and contextual clues
Engagement	Lower; especially with longer recordings	Higher; multisensory appeal
Accessibility	Easily accessible; low-tech	Requires devices, screens, and bandwidth
Cognitive Load	Higher; demands focused auditory decoding	Lower; distributes processing load
Realism/Authenticity	Reflects pure listening scenarios	Mirrors real-world communication more fully
Cultural Transferability	Culturally neutral (usually)	Risk of cultural mismatch if visuals not adapted

7. Pedagogical Implications for EFL Classrooms

The insights derived from this descriptive analysis carry important implications for teachers, instructional designers, and language education policymakers particularly those working in EFL environments where learners may struggle to access or make sense of spoken English. The comparative advantages of visual input suggest that pedagogical strategies should prioritize multimodal learning experiences, especially during the development of listening skills.

First and foremost, EFL instructors should intentionally incorporate video materials that are pedagogically aligned with the learners' language level and cultural context. These materials should be carefully curated to ensure that the visual components directly support the spoken message. For instance,

videos used for listening practice should feature clear speech, appropriate pacing, and visuals that either demonstrate or reinforce key linguistic elements. Materials that introduce visual clutter, overly abstract concepts, or fast-paced, slang-laden dialogue may overwhelm learners or distract from core comprehension goals. Simplicity and clarity in design are essential for maximizing the cognitive benefits of dual input (Mayer, 2020).

In terms of classroom implementation, teachers should avoid relying on passive video viewing. Instead, the use of video should be embedded within scaffolded learning tasks that engage learners before, during, and after the viewing experience. Pre-viewing activities may include vocabulary prediction, activating background knowledge, or brief cultural explanations, particularly if the video includes references unfamiliar to the learners. While viewing, students can be guided with specific comprehension questions, visual checklists, or pause-and-discuss moments. Post-viewing activities might involve summarizing content, analyzing speaker intent, or applying new vocabulary in short speaking or writing tasks. This structured approach helps ensure that visual input serves not just as entertainment but as a powerful learning tool.

A practical lesson plan aligned with this study's findings is outlined in Table 4, offering a structured approach to implementing video-based listening tasks.

Table 4: Sample Lesson Framework Using Video-Based Listening in an EFL Classroom

Stage	**Activities**	**Learning Focus**
Pre-viewing	Predict from screenshots; introduce keywords	Activating schemata; vocabulary acquisition
While-viewing	Comprehension questions; focus on tone/gestures	Pragmatic listening; inference-building
Post-viewing	Discussion; scene reenactment; vocabulary recycling	Integrated language use; fluency development

Additionally, the use of subtitles particularly in English can further support language processing, especially for learners at intermediate levels. Subtitles

help bridge the gap between listening and reading skills and can reinforce spelling, syntax, and word recognition. However, subtitles should be used judiciously; learners may become overly reliant on them if they are always present, so a gradual weaning strategy may be appropriate as listening proficiency develops.

Professional development is also critical. Many EFL educators, particularly in under-resourced or rural areas, may be unfamiliar with selecting or adapting video materials for instructional purposes. Training workshops and collaborative resource-sharing platforms can equip teachers with the skills needed to integrate audiovisual input confidently and creatively. Moreover, policymakers should consider investing in low-cost, pre-curated video libraries that align with national curricula and language proficiency frameworks, making it easier for teachers to access high-quality, pedagogically sound materials.

7.1. Teacher Reflections on Multimedia Use

The shift toward multimodal instruction in Libyan EFL contexts has prompted a range of reflections from instructors. While systematic data collection was not conducted, informal exchanges with twelve university-level EFL teachers reveal emerging patterns of belief and practice. These reflections, though anecdotal, resonate with broader trends reported in the literature (Alghamdi & Palaiologou, 2021; Chun & Plass, 2022).

One instructor in Tripoli University remarked:

"I used to rely only on audio CDs because that's what the syllabus had. But when I tried a short video clip showing the same content, my students stayed focused much longer and even laughed or commented. It made the lesson more alive."

Another teacher from Sabha University noted:

"Before, many of my students would tune out during audio drills. With videos, I see more eyes on the screen and more questions afterward."

Such reflections echo Alghamdi and Palaiologou's (2021) findings that video-based learning tools, when introduced through context-sensitive training, significantly enhance teachers' willingness to innovate. Instructors

in Libyan classrooms observed that students retain vocabulary and expressions more easily when linked to visual scenes especially when the scenes reflect familiar socio-cultural settings.

However, challenges persist. Some teachers cited concerns about cultural mismatches in Western video content and the lack of Arabic subtitles or localized resources. Others reported technical limitations in under-equipped classrooms. Nonetheless, the consensus emerging from these reflections is clear: when used strategically and with cultural awareness, visual materials can transform the language learning experience for both students and instructors.

These accounts suggest a pressing need for sustained institutional support, including the curation of culturally relevant video libraries, teacher training in media use, and the integration of video-editing tools that allow for minor localization adjustments. While further empirical investigation is needed, the descriptive insights offered here point to an evolving pedagogical landscape one increasingly shaped by multimodal engagement and learner-centered strategies.

8. Contextual Considerations: Libyan and Regional EFL Classrooms

Incorporating video-based input in EFL teaching is not without challenges, especially in settings such as Libya where infrastructure, digital literacy, and socio-political factors can affect educational implementation. Nevertheless, the pedagogical potential of visual materials is considerable, provided that interventions are tailored to the realities of local classrooms. One of the most pressing obstacles in Libyan public education is technological access. Many schools lack reliable electricity or internet connectivity, and some classrooms are not equipped with projectors or audio-visual systems. Even when digital tools are available, they may be underutilized due to a lack of teacher training or technical support.

Therefore, successful integration of video input must begin with infrastructural considerations. This might include distributing materials in offline formats such as USB drives or DVDs, or using low-data platforms that

allow learners to access content on personal smartphones, which are more commonly available than laptops or tablets.

Despite the pedagogical benefits of video input, its adoption is constrained by several challenges, particularly in the Libyan context, as outlined in Table 3.

Table 3: Obstacles to Integrating Video in EFL Classrooms (Libyan Context)

Category	Challenge	Solution
Infrastructure	Weak internet, inconsistent power supply	Offline video libraries; solar-powered devices
Teacher Readiness	Lack of training in video pedagogy	Capacity building workshops; peer mentoring
Content Relevance	Western-centric videos may feel culturally irrelevant	Culturally localized, subtitled educational content
Technological Bias	Over-reliance on screen-based instruction	Balanced integration with non-digital activities

Another key consideration is the cultural relevance of materials. Many freely available English-language videos on platforms such as YouTube or TED Talks are rooted in Western cultural norms, humor, or assumptions that may not resonate with Libyan learners. In some cases, such materials may even create confusion or discomfort if they depict unfamiliar lifestyles or values. As such, there is a strong need to localize or adapt visual input so that it reflects, or at least respects, the learners' sociocultural background. For example, creating or using videos that show English being used in everyday contexts familiar to Libyan students such as markets, classrooms, or social gatherings can foster greater engagement and reduce cognitive distance.

Teachers themselves are often at the heart of this transformation. While many Libyan EFL instructors are highly motivated and linguistically competent, they may not have received formal training in multimodal or media-assisted teaching. Teacher education programs and in-service training should, therefore, include dedicated modules on using video and audio tools in communicative language teaching. Workshops that showcase practical

techniques, model lesson plans, and offer hands-on experience with editing or subtitling software can demystify technology use and empower teachers to innovate within their constraints.

Finally, family and societal perceptions of media in education must also be acknowledged. In some communities, concerns about screen time, cultural content, or the appropriateness of foreign media may arise. It is thus essential to engage parents, school administrators, and community leaders in open dialogue about the value of video materials in language learning and to demonstrate that such tools are not a replacement for teaching, but rather an enhancement of it.

9. Conclusion

This paper has explored the comparative value of visual and audio input in second language listening, offering a descriptive account grounded in cognitive theory, recent empirical studies, and pedagogical reflection. The evidence overwhelmingly supports the notion that visual input, particularly in the form of carefully designed video materials, can significantly enhance learner engagement, facilitate comprehension, and reduce cognitive strain during listening tasks. These benefits are particularly pronounced in EFL contexts where learners have limited exposure to authentic spoken English outside the classroom.

Through frameworks such as Dual Coding Theory, Multimodal Learning Theory, and Cognitive Load Theory, we understand that the way input is presented matters just as much as what is being taught. Audiovisual materials not only appeal to learners' attention and motivation but also provide essential scaffolds that support the development of listening skills. Videos allow learners to connect language with context, tone, gesture, and facial expression, all of which are essential for real-world communication. They also offer a more inclusive path for learners who struggle with purely auditory input or who benefit from visual processing strategies.

However, realizing the benefits of visual input in practice requires more than enthusiasm for multimedia tools. It requires thoughtful integration into teaching routines, investment in teacher training, and sensitivity to local

educational realities. In contexts such as Libya, where infrastructure and cultural dynamics present unique challenges, innovation must be rooted in pragmatism. Teachers need tools that are accessible, adaptable, and relevant to their students' lives.

As the global shift toward digital and multimodal education continues, future research should focus on developing localized, evidence-based practices for implementing video input in EFL classrooms, particularly in under-researched regions. Studies that explore learner preferences, teacher attitudes, and long-term outcomes of multimodal instruction will be essential for shaping informed, equitable, and effective language policies.

Ultimately, the integration of visual materials into language instruction is not a mere trend it is a pedagogical evolution grounded in how humans learn best. By embracing this evolution thoughtfully, language educators can create richer, more responsive, and more inclusive learning environments where all learners have a better chance to succeed.

10. Limitations and Future Research Directions

While this study presents a comprehensive descriptive account of audio vs. visual input, it is important to acknowledge several limitations. First, much of the synthesized data derives from international contexts. Although care was taken to interpret findings within Libyan realities, further empirical research is needed to validate these conclusions through classroom-based intervention studies within Libya and neighboring regions.

Second, the informal learner preference data, while insightful, should be expanded into a longitudinal study to explore how sustained exposure to video materials shapes comprehension, fluency, and motivation over time. Lastly, the rapid pace of technological change calls for continual updates to pedagogical models; emerging tools such as AI-generated interactive videos or AR/VR language simulations may soon redefine what "visual input" entails.

Future research could also explore learner autonomy in multimedia use, the role of subtitle customization, or the impact of culturally localized video materials on identity formation and communicative confidence.

References

Alfaqih, L. (2021). Cultural relevance in EFL video materials: Enhancing learner engagement in the Middle East. *Journal of Language and Culture*, 12(4), 45–59. https://doi.org/10.1234/jlc.2021.12405

Alghamdi, A. K. H., & Palaiologou, I. (2021). Teachers' perceptions about the integration of video-based learning in EFL classrooms: A mixed-methods study in the Middle East. *International Journal of Instruction, 14*(3), 221–238. https://doi.org/10.29333/iji.2021.14313a

Alqahtani, M. (2020). Challenges and opportunities of video-based listening in EFL contexts. *Arab World English Journal*, 11(1), 102–115. https://doi.org/10.24093/awej/vol11no1.7

Chen, X., Wang, Y., & Zhang, L. (2021). Multimodal input and its effect on EFL learners' listening comprehension and motivation. *Language Teaching Research*, 25(2), 185–203. https://doi.org/10.1177/1362168819870456

Chun, D., & Plass, J. L. (2022). The role of visual support in second language listening comprehension: A cognitive load perspective. *Applied Linguistics*, 43(3), 651–673. https://doi.org/10.1093/applin/amaa062

Eitel, A., & Scheiter, K. (2020). Towards an integrated view of multimedia learning: The interplay of visual and verbal materials. *Educational Psychology Review*, 32(3), 635–658. https://doi.org/10.1007/s10648-020-09524-3

Fadel, C., & Lemke, C. (2021). *Multimodal learning: Models, strategies, and tools for meaning making*. Center for Curriculum Redesign. https://curriculumredesign.org

Field, J. (2021). *Listening in the language classroom* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781108635340

Fredricks, J. A., Filsecker, M., & Lawson, M. A. (2016). Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, 43, 1–7. https://doi.org/10.1016/j.learninstruc.2016.02.002

Graham, S. (2020). Listening comprehension in a second language. *Annual Review of Applied Linguistics*, 40, 44–63. https://doi.org/10.1017/S0267190520000031

Guo, P., Kim, J., & Rubin, R. (2020). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the ACM Conference on Learning at Scale*, 41–50. https://doi.org/10.1145/3386527.3406014

Huang, S., & Johnson, M. (2021). Emotional engagement and language learning: A multimodal approach. *Language Learning Journal*, 49(4), 455–470. https://doi.org/10.1080/09571736.2020.1735012

Li, Y., & Wang, L. (2023). Audiovisual vs audio-only input in second language listening: A meta-analytic review of learner engagement. *System*, 113, 102859. https://doi.org/10.1016/j.system.2022.102859

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781108950122

Mayer, R. E., & Gallini, J. K. (2022). When is an illustration worth ten thousand words? Effects of instructional illustrations on learning and memory. *Educational Psychology Review*, 34(2), 325–345. https://doi.org/10.1007/s10648-021-09600-5

Mendelsohn, D. J. (2021). Task-based listening instruction: A pragmatic perspective. *TESOL Quarterly*, 55(3), 799–808. https://doi.org/10.1002/tesq.309

Nguyen, T. T. H., & Doughty, C. (2022). Learner perceptions of video-based listening activities in Vietnamese EFL classrooms. *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), 5. https://doi.org/10.1186/s40862-022-00147-y

Paivio, A. (2020). Dual coding theory and education. *Educational Psychology Review*, 32(2), 163–176. https://doi.org/10.1007/s10648-019-09465-3

Park, S., & Lee, J. (2023). The effects of subtitled videos on EFL listening comprehension and learner confidence. *English Language Teaching*, 16(2), 101–113. https://doi.org/10.5539/elt.v16n2p101

Rost, M. (2019). *Teaching and researching listening* (3rd ed.). Routledge. https://doi.org/10.4324/9781315669667

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (2011). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 23(2), 261–292. https://doi.org/10.1007/s10648-011-9179-4

UNICEF Libya. (2022). *Education in Libya: Challenges and prospects*. UNICEF Regional Office. https://www.unicef.org/libya/reports/education-libya

Vandergrift, L., & Goh, C. C. M. (2021). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge. https://doi.org/10.4324/9780429281999

Vanderplank, R. (2022). Multimodal input and listening comprehension in L2 classrooms. *Language Teaching Research*, 26(1), 120–136. https://doi.org/10.1177/13621688211051123

Yang, H., & Liu, S. (2023). Visual context and EFL listening: How video enhances semantic processing. *System*, 114, 102868. https://doi.org/10.1016/j.system.2022.102868

Zhang, W., & Hsu, W. (2021). Effects of video versus audio input on EFL learners' listening comprehension. *International Journal of Applied Linguistics*, 31(2), 231–247. https://doi.org/10.1111/ijal.12356