
| RESEARCH ARTICLE

Sonic Scenography and the 3D Binaural Turn: A Multimodal Framework for Enhancing EFL Acquisition Among Completely Visually Challenged Moroccan Students

Meriem Daoui ¹✉ and **Sana Sakale** ²

^{1,2} Faculty of Languages, Letters and Arts, Ibn Tofail University, Kenitra, Morocco

Corresponding Author: MERIEM DAOUI, **E-mail:** meriem.daoui1990@gmail.com

| ABSTRACT

This research addresses the pedagogical vacuum created by ocularcentric English Language Teaching (ELT) materials for completely visually challenged students in the Moroccan context. In the absence of visual imagery, this study proposes a transformative "Sonic Scenography" model, utilizing binaural 3D soundscapes to scaffold lower-order linguistic skills. Employing a mixed-methods approach, the study examines an online cohort of 15 students with a complete visual challenge across various Moroccan cities. By integrating elements of theatrical sound design and cognitive psychology, the research explores how spatialized audio can serve as a functional surrogate for visual context. Quantitative data were analyzed using Statistical Package for the Social Sciences (SPSS) software. Findings reveal that 3D acoustic environments significantly enhance vocabulary retention and phonological awareness by providing a "spatial anchor" for new language data. The results indicate that shifting from "flat" audio to immersive 3D scenography reduces cognitive load and fosters academic autonomy for completely visually challenged students.

| KEYWORDS

Sonic Scenography; Binaural 3D Audio; Completely Visually Challenged Students; Cognitive Load Theory; EFL Vocabulary Retention

| ARTICLE INFORMATION

ACCEPTED: 01 February 2026

PUBLISHED: 20 February 2026

DOI: 10.32996/ijels.2026.8.1.2

1. Introduction

1.1. Background of the Study

The dominance of vision in Western epistemology, described by Jay (1993) as ocularcentrism, has historically framed the "spectatorial nature" of knowledge acquisition. This hegemony is etched into the very etymology of academic discourse; as Styhre (2010) notes, the term "theory" stems from the Greek *theoria*, meaning to look at or behold. In the realm of English Language Teaching (ELT), this has resulted in a "spectatorial epistemology" where textbooks and digital media prioritize visual cues as the primary conduit for meaning (Chmielecki, 2025). In Morocco, the educational landscape has undergone significant shifts through the "Strategic Vision 2015-2030" and the 2019 "no child left behind" program, which aim to transition from a medical model of disability to a social model of inclusion (Naciri, 2022). However, for completely visually challenged students, the current curriculum remains largely image-dependent, creating a structural barrier that necessitates a radical sensory shift. The emergence of "Sonic Scenography"—the art of designing spatialized auditory environments—offers a decolonizing alternative to the ocularcentric status quo by leveraging 3D binaural technology to build a "theater of the mind" (Verma, 2012).

1.2. Problem Statement

There exists a critical pedagogical gap in the development of lower-order linguistic skills—specifically vocabulary recognition, phonetic discrimination, and basic syntax—when traditional visual scaffolds are absent. Current Moroccan ELT materials are designed for sighted learners, treating the "absence of the image" for completely visually challenged students as a deficit rather

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (<https://creativecommons.org/licenses/by/4.0/>). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

than an opportunity for alternative sensory engagement. This reliance on 2D, mono-aural audio aids often results in a "contextual void," where learners with a complete visual challenge struggle to anchor new vocabulary to spatial or situational concepts (Asebriy et al., 2018). Without immersive contextualization, the cognitive effort required to process English grammar increases, leading to higher levels of frustration and lower academic performance.

1.3. Research Objectives

To evaluate the effectiveness of 3D binaural soundscapes as a replacement for visual imagery in enhancing vocabulary retention among completely visually challenged students.

To design a "Sonic Scenography" framework that utilizes theatrical sound design (Foley art) to scaffold phonetic discrimination skills for students with a complete visual challenge.

To analyze the impact of immersive spatialized 3D audio on reducing the cognitive load associated with basic syntactic comprehension in an online environment.

1.4. Research Questions

1. How does the transition from 2D "flat" audio to 3D "spatialized" sound affect the acquisition of vocabulary among completely visually challenged students?
2. In what ways can theatrical sound design (Foley art) provide the contextual scaffolding necessary for phonetic discrimination and phonological awareness?
3. To what extent does the use of spatialized 3D audio reduce the cognitive load associated with basic syntactic comprehension for Moroccan students with a complete visual challenge?

1.5. Significance of the Study

This study is of paramount significance as it aligns with the "Digital Morocco 2030" strategy, which emphasizes "Inclusive Usage" and the digitalization of services for all citizens (MMSP, 2024). It provides an empirical foundation for Moroccan curriculum designers to move beyond "image-heavy" pedagogy. Furthermore, by validating the unique auditory-spatial intelligence of completely visually challenged students, the study promotes a more equitable and democratic educational environment in North Africa.

1.6. Key Concepts

- **Sonic Scenography:** The intersection of theater arts and linguistics to create a 360-degree, immersive auditory learning environment (Curtin & Roesner, 2015).
- **Binaural 3D Audio:** A recording and playback technique that uses Head-Related Transfer Functions (HRTFs) to mimic how human ears locate sounds in three-dimensional space (Shimizu et al., 2022).
- **Students with a Complete Visual Challenge:** Learners who lack functional visual perception and rely on auditory and tactile modalities for information processing.
- **Lower-Order Skills:** Foundational linguistic competencies including vocabulary acquisition, phonetic awareness, and simple syntactic structures.

2. Theoretical Framework

The study is grounded in three primary theoretical pillars that explain the intersection of sensory perception and language acquisition:

2.1. Embodied Cognition Barsalou (2008) and Zwaan (2004) posit that language meaning is grounded in motor-perceptual experiences. When learners hear an action verb, the brain activates neural pathways associated with the physical act. For completely visually challenged students, auditory-spatial cues serve as the primary grounding mechanism for linguistic concepts. Meaning is not an abstract symbol but an "immersive experience" within the body (Aziz-Zadeh & Damasio, 2008).

2.2. The Theatre of the Mind Neil Verma (2012) explores how acousmatic sound (sound without a visible source) triggers internal mental imagery. This "excavational listening" allows students to construct structured "soundscapes" that replace the missing visual image. For students with a complete visual challenge, this internal theater becomes the primary site for "auditory sense-making".

2.3. Cognitive Load Theory (CLT) Sweller (2011) suggests that learning is inhibited if the working memory is overwhelmed by extraneous information. Traditional audio often requires high "mental demand" to decipher context. Spatialized audio reduces this load by providing "real-world" cues that the brain processes instinctively, freeing up resources for linguistic schema construction (Sweller et al., 2011).

3. Conceptual Framework

The proposed conceptual model is the Sonic Proscenium. In this model, "Space" is treated as a grammatical variable rather than just an environment. By placing specific parts of speech at different 3D coordinates (e.g., Subjects to the left, Verbs in the center, Objects to the right), the framework uses spatial sequence to represent syntax. This "Acoustic Scenography" transforms the virtual classroom into a 360-degree stage where the student is the central actor. The model operates on the "Spatial Peg" hypothesis: linguistic labels are "anchored" to specific 3D coordinates in the learner's mental theater, facilitating faster retrieval and deeper encoding (Verma, 2012).

4. Literature Review

4.1. Ocularcentrism and Inclusive Applied Linguistics

In the investigation of visual dominance, Jay (1993) aimed to critically examine the historical reverence for vision in Western philosophy through his seminal work, *Downcast Eyes*. The study found that Western culture has been governed by an ocularcentric paradigm where sight is equated with truth, often marginalizing alternative sensory modes. However, Jay (1993) missed an analysis of how this "hegemony of the eye" specifically compromises Second Language Acquisition (SLA) for completely visually challenged students. This research helps the present study by establishing the philosophical foundation for the "epistemic barrier" that ocularcentric ELT materials present to learners who process meaning through non-visual channels.

Additionally, Born (2013) aimed to analyze how sound and music create "relational spaces" that define social and sensory hierarchies. The research found that sound is not merely an aesthetic background but a primary site for meaning-making and the construction of identity. Born (2013) missed the specific potential of these "sonic territories" as formal scaffolds for grammatical comprehension in EFL. This investigation helps the current research by providing the theoretical justification for treating the auditory field as a legitimate academic space for students with a visual challenge.

4.2. Psycholinguistics of Visual Loss

Addressing the neuro-cognitive aspects, Voss et al. (2004) aimed to investigate sound localization abilities and identify the neural substrates used by individuals with a complete visual challenge. They found that visual loss leads to superior azimuthal localization and the recruitment of the occipital cortex—normally used for vision—for auditory spatial processing. The researchers missed a connection between this cortical plasticity and the retention of linguistic schemas like syntax. This study helps the current investigation by proving that completely visually challenged students possess a unique "auditory-spatial intelligence" that can be optimized through 3D scenography.

Further evidence is provided by Roder et al. (1999), who aimed to study how the brain processes auditory space following visual loss. They found that crossmodal recruitment allows the brain to handle spatial cues with higher precision. However, they missed evaluating how this enhanced precision can act as a surrogate for visual picture dictionaries in a classroom. This helps the current study by supporting the "Spatial Peg" hypothesis, suggesting that spatialized sound can act as a definitive grounding mechanism for new vocabulary.

4.3. Educational Technology and Inclusive Performance

In the context of digital success, McNicholl et al. (2021) conducted a systematic review to evaluate the impact of assistive technology use on students with disabilities in higher education. Their findings showed that tools providing high-quality auditory feedback significantly improved task efficiency and academic performance, though barriers to integration remain. The study missed an exploration of 3D binaural soundscapes as a sophisticated evolution of standard screen-reading software. This research helps the present study by validating that students with a visual challenge experience measurable linguistic gains when digital materials are tailored to their sensory profiles.

Finally, Asebriy et al. (2018) aimed to enhance the accessibility of Moroccan educational environments for learners with a visual challenge through specialized ICT solutions. The results demonstrated that assistive technologies could empower these students to perform academic activities previously deemed impossible, though curricular materials often lacked relevant grounding. They missed a focus on "Sonic Scenography" or foley-based audio to fill the "contextual void" in the English curriculum. This helps the

current research by demonstrating the readiness of Moroccan learners to adopt high-fidelity audio frameworks to bypass standard ocularcentric barriers.

5. Methodology

5.1. Research Design

This study utilized a quasi-experimental, mixed-methods approach conducted entirely online through synchronous platforms optimized for screen readers and high-fidelity audio transmission.

5.2. Participant Selection

A purposive sample of 15 completely visually challenged students was selected from diverse Moroccan cities (Rabat, Casablanca, Fez, and Agadir). All participants were adults enrolled in a specialized online English reinforcement program.

5.3. The Intervention: Sonic Scenography

The intervention consisted of creating "3D Audio-Scenes" using binaural recording techniques:

- Theatrical Scenography: Foley sound effects (e.g., a kettle boiling to the right to represent the noun "kettle") replaced visual flashcards.
- Syntactic Mapping: Sentences were presented in 3D space (Subject: 45° left; Object: 45° right).
- Action Verb Priming: Students listened to 18 Hz Beta binaural beats for 10 minutes prior to lessons to enhance syntactic interpretation.

5.4. Data Collection and Analysis

Quantitative data were gathered via pre-tests and post-tests. The SPSS software (version 26.0) was used to conduct a Paired Samples T-Test to determine the significance of improvements. Qualitative data were collected through semi-structured audio interviews and "audio-narrative diaries."

5.5. Ethical Considerations

Informed consent was obtained via auditory means (recorded verbal consent). All research tools were designed for screen-reader compatibility (WCAG 2.2 standards). Data privacy was maintained through anonymized digital identities.

6. Findings

The intervention yielded the following results based on SPSS analysis:

6.1. Linguistic Performance Gains

Analysis revealed a highly significant improvement in performance ($p < .05$). The mean score for vocabulary retention rose from 48% (pre-test) to 82% (post-test), with a large effect size of $d = 0.832$. This improvement was particularly noticeable in the retention of concrete nouns and transitive verbs. For instance, when a professional voice actress recorded the word "fetch" as a sound movement moving from side to side around a dummy head microphone, students were able to recall the word with 95% accuracy one week later. This suggests that the binaural recording technology alone, by providing a unique episodic memory of movement, acts as a powerful mnemonic device for students with a visual challenge, effectively replacing the visual picture dictionary.

6.2. Syntactic Scaffolding Success

Students demonstrated a 36% increase in accurate Subject-Verb-Object (SVO) recognition. The use of Foley sound sequences (e.g., footsteps followed by a door slam) helped 94% of students identify transitive action verbs correctly without any visual aids. By placing the "Subject" (the source of the sound) at a 45° left coordinate and the "Object" (the result of the sound) at a 45° right coordinate, the students utilized their azimuthal localization skills to "feel" the syntax. The teacher noted that "the sound became a physical map of the sentence," allowing students to manipulate complex syntactic structures mentally by tracking the movement of sounds through their internal "theater of the mind."

6.3. Cognitive Load Reduction

According to the NASA-TLX dimensions, "Mental Demand" decreased by 62% for completely visually challenged students when using 3D scenography compared to standard flat audio. "Frustration Levels" also saw a marked reduction of 69%. The qualitative data from audio-narrative diaries revealed that students felt a sense of "presence" in the virtual 3D classroom. One student remarked that in standard audio, they have to "work hard to guess where the teacher is," but in the 3D soundscape, the spatial cues were processed "instinctively." This reduction in cognitive load indicates that the brain's "where" system handles the spatializing work, freeing up significant working memory resources for the actual linguistic task of decoding English grammar.

7. Discussion

The success of the Sonic Scenography model confirms that the "absence of the image" is not a cognitive deficit but a call for multimodal realignment. The 35% increase in complex syntactic comprehension suggests that spatial audio provides a "pseudo-visual sensation" (Susanto & Nanda, 2018) that allows the brain's "where" system to assist the "what" system in language processing. This phenomenon aligns with the Embodied Cognition hypothesis, where meaning is not found in abstract symbols but in the body's simulation of physical reality. By placing sounds in a 360-degree environment, we allow completely visually challenged students to utilize their superior auditory-spatial localization skills—identified in the Voss et al. (2004) study—as a direct replacement for the visual prompts typically found in Moroccan ELT textbooks.

Furthermore, the significant reduction in cognitive load supports Sweller's (2011) theory that immersive multimodal cues allow the brain to focus on schema construction rather than environment-mapping. In traditional mono-aural instruction, students must use high "mental effort" to construct a context for the words they hear. In contrast, the Sonic Proscenium provides a "spatial peg" (Verma, 2012) where linguistic labels are anchored to specific 3D coordinates. This reduces the "contextual void" and the resulting frustration described in the problem statement (Asebriy et al., 2018). For completely visually challenged students in Morocco, this technological shift fosters a sense of "presence" and agency, transforming them from passive listeners into active participants who can navigate the linguistic landscape with confidence. By leveraging the "sonic turn," we validate the unique auditory-spatial intelligence these students bring to the academic sphere, fulfilling the "Digital Morocco 2030" goal of inclusive and equitable usage.

8. Recommendations

1) 8.1. Curricular Reform

The Moroccan Ministry of National Education should formally incorporate "audio-first" ELT materials that prioritize the auditory-spatial intelligence of students with a complete visual challenge. This involves a shift away from "Look at the picture" prompts toward high-fidelity binaural soundscapes that ground vocabulary in spatial reality.

8.2. Technical Training

The "GENIE Program" and "Morocco DigiSchool" should include workshops for teachers on "Sonic Pedagogy" and basic Foley engineering. Training should empower educators to create spatialized "sound sequences" for digital instruction, utilizing dummy head microphones and spatial audio software to build immersive virtual classrooms.

8.3. Policy Advocacy

Policymakers must ensure that Moroccan digital platforms support high-fidelity 3D audio transmission. This includes the provision of high-quality headphones as part of the "Digital 2030" infrastructure, ensuring that every student with a visual challenge has the hardware necessary to access immersive educational content.

9. Conclusion

This research proves that the absence of the image is not a deficit, but an opportunity to refine the "Sonic Proscenium" of the human mind. By merging theater arts with Binaural Audio, we provide a profound framework for the Moroccan ELT field. Space, sound, and movement are the primary languages of the completely visually challenged student. Embracing the "sonic turn" ensures that every Moroccan learner has an equal opportunity to master English and thrive in the global digital economy.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] Asebriy, Z., Anyuru, A. & Raghay, S. (2018). Towards an accessible educational environment for visually challenged learners: An ICT-based approach. *Journal of Mobile Multimedia*, 14(3), 291-314.
- [2] Barsalou, L. W. (2008). Grounded cognition. *The Annual Review of Psychology*, 59, 617–645.
- [3] Curtin, A. & Roesner, D. (2015). Sounding out scenography: The multimodal turn in theatre. *Theatre and Performance Design*, 1(1), 107-125.
- [4] Jay, M. (1993). *Downcast Eyes: The Denigration of Vision in Twentieth-Century French Thought*. University of California Press.
- [5] McNicholl, A., Casey, H., Desmond, D. & Gallagher, P. (2021). The impact of assistive technology use for students with disabilities in higher education: A systematic review. *Disability and Rehabilitation: Assistive Technology*, 16(2), 130-147.
- [6] MMSP. (2024). *Digital Morocco 2030 Strategy*. Ministry of Digital Transition and Administration Reform.
- [7] Naciri, M. (2022). Morocco's inclusive education program through the lens of ethnography. *International Journal of Linguistics, Literature and Translation*, 5(6), 34-45.
- [8] Styhre, A. (2010). *The Epistemology of the Eye: Inscriptions and Organizations*. Ashgate Publishing.
- [9] Verma, N. (2012). *Theater of the Mind: Imagination, Aesthetics, and American Radio Drama*. University of Chicago Press.
- [10] Voss, P., Lassonde, M., & Lepore, F. (2004). Early-blind human subjects localize sound sources better than sighted subjects. *Neuron*, 41(5), 695-697.
- [11] Zwaan, R. A. (2004). The immersed experienter: Toward a perceptual theory of language comprehension. *The Psychology of Learning and Motivation*, 44, 35-62.