

Research Paper

SonicLens: Visual Metaphors and Explainable AI for Formative Feedback in Sound Design Education

*Razif Mohamed

The Design School, Faculty of Innovation and Technology, Taylor's University, Malaysia

© The Author(s) 2026. This article is published with open access by Taylor's Press.

Abstract: SonicLens is an artificial intelligence application designed to assist design students as they encounter the complexities of sound design. Within design education, critique is widely recognized as the signature pedagogy that shapes how students reflect, iterate, and refine their creative work. However, many novice learners find the technical vocabulary of audio engineering difficult to interpret, while instructors often face challenges in providing the frequent, high-resolution feedback required in large studio settings. This logistical gap limits the ability of students to refine their work through iteration, which is fundamental to the creative learning process. SonicLens addresses this by merging automated audio analysis with explainable guidance, translating abstract audio concepts into visual design metaphors grounded in cross-modal perception research. These metaphors connect unfamiliar auditory principles such as dynamic range, equalization, and compression to familiar visual practices like contrast, color balance, and exposure control, making technical ideas more approachable and encouraging students to reflect with greater confidence. The system utilizes Explainable AI (XAI) and the XAIxArts framework to ensure that machine-generated feedback remains transparent and pedagogically sound. A modular six-stage prompt-chain architecture governs the AI's feedback generation, ensuring consistency, transparency, and alignment with the pedagogical objectives of the design studio. By providing AI-guided critique alongside human input from instructors and peers, SonicLens extends the reach of formative feedback beyond the physical classroom while preserving the relational nature of design critique. This paper presents the pedagogical grounding, system design, and technical architecture of SonicLens, situating it within broader developments in AI-augmented critique. Formal empirical evaluation is planned as future work.

Keywords: Soniclens, explainable AI, sound design education, visual metaphors, formative feedback, XAIxArts

Suggested citation: Mohamed. R. (2026). SonicLens: Visual metaphors and explainable AI for formative feedback in sound design education. *Asia-Pacific Journal of Futures in Education and Society*, 5(3), 107–131.

*Correspondence: Razif Mohamed, Taylor's University. Email: razif.mohamed@taylors.edu.my

Introduction

Critique is widely recognized as the signature pedagogy of design education. Building on Schön's (1983) notion of the reflective practitioner, design critique provides a structured environment where students reflect on their creative choices, articulate their intent, and refine their work through social interaction (Shulman, 2005). In the domain of sound design, however, critique is often hindered by the temporal and invisible nature of the medium. Unlike visual disciplines, where feedback can be annotated directly onto a static canvas, sound-based critique requires attentive listening and time to analyse performance over a specific duration. These conditions make it difficult for instructors to offer the immediate and frequent responses necessary for effective iterative learning, particularly in hybrid or large-enrolment classrooms.

A significant barrier for students in a design-centric environment is the translation gap between visual intuition and auditory physics. Many design students struggle to grasp technical terms such as dynamic range, spectral balance, and spatial imaging because these concepts lack the immediate visual reference points they rely on in their primary training. While a student can instantly recognize a cluttered layout or a lack of visual contrast, the equivalent issues in a soundscape are often perceived but difficult to name. This lack of a shared vocabulary between the student and the technical demands of the medium creates a friction that limits creative confidence and prevents deep engagement with technical feedback.

Recent advancements in generative and explainable AI offer new pathways to address this friction by reframing critique as a collaborative dialogue between human and machine. AI-driven feedback systems can now evaluate complex media such as sound, providing insights that support technical accuracy and conceptual exploration (Duan et al., 2024). Research suggests that integrating artificial intelligence into design-based learning can enhance reflective thinking and improve the ability of students to analyse and iterate on their work (Saritepeci & Durak, 2024). Within this context, SonicLens is designed to assist students in understanding the performance of their compositions across measurable parameters while helping them interpret the narrative dimensions of their sound design.

This paper examines the design and technical logic of SonicLens as a pedagogical innovation. It explores how automated feedback can support reflection and iteration without displacing the essential role of the human instructor. The contribution of this work is presented as a design and system account rather than an empirical study. By establishing the technical architecture and theoretical framework of the system, this paper provides the foundation for subsequent classroom deployment and formal evaluation.

Literature Review

The practice of critique is fundamental to creative education and is recognized as the signature pedagogy that shapes how design students think and act within their disciplines (Gray & Howard, 2015; Shulman, 2005). Critique provides a structured environment where disciplinary knowledge is socially constructed through a dialogic exchange between student and mentor. In this constructivist model, knowledge is not merely transmitted but co-created through iterative reflection (Crowther, 2013). For sound design education, however, this dialogue is often impeded by the technical complexity of the medium, which lacks the immediate visual affordances found in other design fields. Even audio professionals struggle to articulate experiential qualities of sound, with terms like “juicy” or “punchy” serving as imprecise descriptors that resist standardization (Hicks et al., 2024).

Despite the importance of critique, traditional studio environments often encounter challenges related to performance anxiety (McCarthy, 2011). When students are forced to process technical feedback in real-time without a shared conceptual vocabulary, the critique can shift from a creative dialogue to a one-way technical correction. To address this, Lim et al. (2023) suggest that AI systems can function as “relational artifacts” that facilitate a safer, asynchronous space for reflection. This shift potentially enables “assessment for learning,” where the focus remains on the iterative refinement of the work (Memarian & Doleck, 2024).

The value of peer interaction extends beyond instructor-led critique. Yong and Lee (2026) found that structured peer learning systems support collaborative knowledge construction, particularly when scaffolded through intentional pedagogical design. This suggests that AI feedback systems, when integrated alongside peer commentary, may create richer learning environments than either approach alone.

The Decoding Gap: Experts vs. Novices

A primary challenge in sound design pedagogy is the reliance on visualization tools such as spectrograms and waveform displays. While these tools provide objective data, they are designed for “expert listening”, a state where the user can already decode visual patterns into auditory meaning (Giannakis, 2006; Magnusson, 2017). For a novice designer, a spectral peak or a dynamic range shift remains an abstract technical phenomenon. Research suggests that novices often lack the mental models required to translate raw visual data into actionable creative decisions (Elmosnino, 2022). Pirhonen (2014) argues that metaphor can serve as a focal concept in sound design, offering interpretative meaning that bridges this kind of gap. SonicLens draws on this principle to address the “decoding gap” between novice and expert listeners. By mapping auditory data to visual principles, the system is intended to scaffold the student’s transition from intuitive visual reasoning to technical auditory proficiency.

Theoretical Foundations of Cross-Modal Mapping

The use of visual metaphors in SonicLens is grounded in cross-modal perception research, which suggests that associations between visual shapes and sound qualities are robust and intuitive (Liu et al., 2025; Walker, 1987). Pedagogy based on the sonification of visual structures seeks to allow students to grasp abstract auditory concepts by drawing on their existing design literacy (Joo, 2022; Somers, 1998).

These mappings are designed to utilize several intersecting theoretical frameworks to bridge the auditory and visual domains. This includes the application of “embodied schemas” to explain volume and spatiality (Lemaitre et al., 2017), the use of “cross-modal perception” to map frequency to colour and shape (Ward et al., 2006), and “structural translation” to visualize abstract mix hierarchies (Somers, 1998). Furthermore, the system relies on “graphic representation” and “synesthetic design” principles with the goal of ensuring that technical feedback aligns with the gestalt continuity of a narrative (Fahlenbrach, 2008). Table 1 outlines these theoretical foundations and the intended learning outcomes for the metaphor mappings employed in the SonicLens system.

Table 1. Theoretical foundations for SonicLens audio-to-visual metaphor mappings

Audio Concept	Visual Metaphor	Theoretical / Research Basis	Intended Learning Outcome
Dynamic Range	Contrast in a layout	Embodied Schema: Spatial and visual schemas are robust tools for explaining volume (Lemaitre et al., 2017).	Students learn that impact depends on variation, not just volume.
Spectral Balance	Colour palette harmony	Cross-modal Perception: Users instinctively associate specific visual shapes and colours with sound quality (Ward et al., 2006).	Excessive frequency in one band feels like oversaturation.
Mix Depth	Visual hierarchy	Structural Translation: Students grasp abstract structures when translated into visual compositions (Somers, 1998).	Foreground and background placement must be intentional.
Timing & Groove	Alignment and grid	Graphic Representation: Visual scores help novices manipulate and understand sound structures (Giannakis, 2006).	Consistent timing maintains clarity and flow.
Space & Ambience	Negative space	Synesthetic Design: Designers use visual language to specify technical space and timbre (Fahlenbrach, 2008).	Silence and reverb shape perception and rhythm.

Table 1. (con't)

Audio Concept	Visual Metaphor	Theoretical / Research Basis	Intended Learning Outcome
Transitions	Gestalt continuity	Audio-visual Metaphor: Visual metaphors guide technical decisions for narrative coherence (Fahlenbrach, 2008).	Smooth progression supports narrative coherence.

AI-Augmented Critique in Design Education

Recent work on AI-augmented critique has largely focused on visual design contexts. Tools such as UICrit demonstrate how artificial intelligence can support heuristic feedback based on established design principles (Duan et al., 2024). However, these systems are primarily oriented toward static imagery and user interface evaluation, leaving the temporal and non-visual characteristics of sound design underexplored. SonicLens extends this line of work into the auditory domain by addressing sound as a time-based and perceptually invisible medium. By combining persona-based critique (Shulman, 2005) with technical audio metrics, the system adopts an explainable AI (XAI) approach that emphasizes pedagogical transparency and reflective understanding, rather than automated correction alone (Bryan-Kinns et al., 2023).

Beyond visual design, the integration of digital applications in performance-based creative education has shown promise in supporting student learning. Zhu and Krish (2026) demonstrated how digital tools in music education can transform pedagogical practices in tertiary institutions, suggesting that technology-mediated feedback systems may have broad applicability across creative disciplines.

Limitations of Metaphorical Scaffolding

While visual metaphors facilitate initial comprehension, they are not a universal solution. Cultural background and prior training influence how individuals interpret cross-modal mappings (Walker, 1987). Furthermore, an overreliance on metaphors may lead to the oversimplification of complex sonic phenomena (Pirhonen, 2014). To mitigate these risks, SonicLens is designed as a collaborative tool that invites human-in-the-loop interaction, ensuring that AI-generated metaphors serve as a provocation for further critical engagement with the instructor.

Methodology

SonicLens is developed as a web-based workspace designed to facilitate a multi-source critique process. The interface is intentionally minimalist to minimize cognitive load, allowing students to focus on their creative intent rather than technical navigation.

When students enter the system, they are prompted to upload their audio files and provide a narrative description of their sound design objectives (Figure 1).

Figure 1. SonicLens dashboard for project setup. Students upload an audio file, specify project type and storyline, then initiate the AI analysis process.

Overview and Roles

The system is designed to support three distinct roles: the student author, peers, and the instructor. All participants interact within a shared environment that centres on the audio waveform. This structure is intended to ensure that feedback is not isolated but exists as part of a continuous pedagogical dialogue.

Workflow and Architecture

The architecture of SonicLens follows a sequential workflow from the initial upload to the final reflective stage. Each phase of the process is designed to combine automated

insights with human interpretation, ensuring that the critique remains grounded in the student's narrative goals.

1. **Upload:** Students submit an audio file alongside a brief statement of design intent.
2. **Pre-processing:** The system performs initial file validation and extracts technical metadata.
3. **AI Analysis:** A multimodal analysis is performed to evaluate loudness, spectral balance, and spatial imaging.
4. **Feedback Assembly:** The system organizes the results into a structured schema for delivery.
5. **Human-in-the-loop Interface:** Students engage with the feedback through a waveform view, while instructors and peers add timestamped annotations.
6. **Persistence and Sharing:** All interactions and revisions are stored to support ongoing self-assessment. The complete workflow is illustrated in Figure 2.

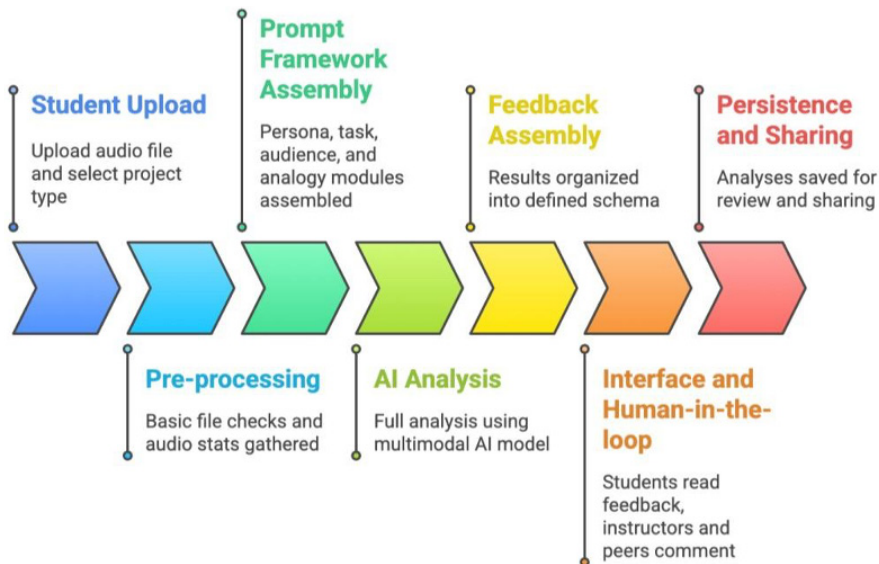


Figure 2. SonicLens workflow illustrating the process from student upload to shared feedback

Technical Implementation

To ensure technical accuracy and prevent model hallucination, SonicLens utilizes a two-stage analysis pipeline grounded in a tool-calling architecture. The first stage extracts quantitative metrics using a specialized Genkit tool called “getTechnicalAudioDetails”. This tool invokes a server-side service that decodes

compressed audio formats, such as MP3, into raw Pulse Code Modulation (PCM) data using the audio-decode library.

Once decoded, the “pcm-util” library performs precise mathematical calculations on the raw data to extract six core technical metrics: Duration, Channel Count, Sample Rate, Bit Depth, average Root Mean Square (RMS) loudness, and Peak levels. This structured data is returned as a JSON object, providing a factual foundation for the second stage.

In the second stage, Gemini 2.5 Flash performs a qualitative analysis. By receiving both the raw audio file and the extracted technical metrics, the model is designed to ground its observations in precise measurements. This dual approach ensures that qualitative statements, like the suggestion that audio is nearing distortion, are anchored in specific peak values and RMS trends. This avoids generic or irrelevant suggestions.

The Prompt-Chain Process

The feedback generation is governed by a modular prompt-chain consisting of modules P1 to P6. This structure ensures that AI maintains a consistent pedagogical persona and follows a logical reasoning sequence. The chain is designed to operationalize the theoretical frameworks established in the literature review.

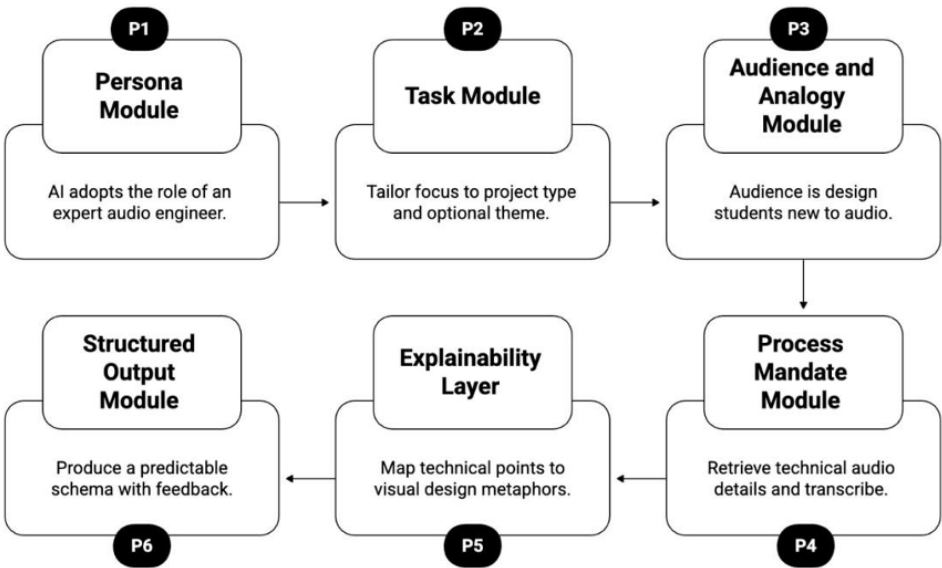


Figure 3. SonicLens prompt-chain process (P1–P6) showing the modular structure used for generating explainable and context-aware AI feedback.

The six modules, labelled P1 to P6 in Figure 3, represent the sequential logic used to generate explainable feedback. The AI feedback process in SonicLens operates through a six-part prompt-chain design, each module defining a distinct function within the reasoning sequence.

P1. Persona Module

AI adopts the role of an expert audio engineer who provides critique in the style of studio review.

Example prompt structure:

You are an expert audio engineer providing feedback on sound design and audio quality. Your analysis is for design students who are new to audio but familiar with visual design principles.

P2. Task Module

Tailors the feedback focus to the project type and optional theme.

Examples: dialogue → clarity and intelligibility;

sound effects → impact and texture;

soundscapes → immersion and atmosphere.

Example prompt structure:

*Your analysis MUST be tailored to the specific project type provided by the user.
Project Type: "Soundscape Design"
Creative Theme: "A calming walk through a fantasy forest at night"*

P3. Audience and Analogy Module

Defines the target audience as design students new to audio. A guiding rule is applied: state the audio term first, then explain it through a visual design metaphor.

Example prompt structure:

*CRITICAL Feedback Structure:
To make your feedback more relatable, you MUST use analogies from graphic design or photography. For every point of feedback, follow this three-part structure:*

- 1. State the Concept (e.g., 'Dynamic Range')*
- 2. Give Actionable Advice*
- 3. Provide the Analogy*

Here is your knowledge base of analogies:

Audio Concept → Visual Metaphor

Equalisation (EQ) → Colour Balance / Grading

Compression → Exposure Control

Dynamic Range → Contrast in Layout

P4. Process Mandate Module

Retrieves technical audio details and generates an internal transcription. Grounds all feedback timestamps in the transcription and ensures coverage of the entire track.

Example prompt structure:

Analysis Steps:

- 1. Ground Feedback in Audio Events: Your timestamped feedback must accurately correspond to events in the audio.*
- 2. Populate all fields in the 'technicalOverview' object using the technical data provided.*

Technical Data:

```
{
  "duration": 120.5,
  "channels": 2,
  "sampleRate": 44100,
  "bitDepth": 16,
  "rms": [0.45, 0.46],
  "peak": [0.95, 0.92]
}
```

P5. Explainability Layer (XAIxArts)

Maps each technical observation to a visual design metaphor, translating complex sound analysis into familiar design language

Example prompt structure:

For each piece of feedback you provide, you must explain your reasoning by referencing the specific audio data that led to your observation.

Example explainability format: "At 0:45-1:10, the RMS levels drop significantly (from 0.45 to 0.28) while the dialogue continues. This creates a clarity issue - like text that's too light against its background in visual design. The viewer has to strain to read it."

Always connect your observations to: - Specific timestamp ranges - Measurable data (RMS, peak, spectral characteristics) - The visual design metaphor from your knowledge base

P6. Structured Output Module

Produces a predictable schema for feedback delivery:

- Technical overview with one-sentence summary
- Overall feedback
- Audio quality feedback (loudness, distortion, dynamic range, clarity, balance)
- Sound design feedback (timestamped issues, general suggestions).

Example prompt structure:

You must return your analysis as a valid JSON object following this exact schema:

```
{
  "technicalOverview": {
    "duration": [number],
    "channels": [number],
    "sampleRate": [number],
    "bitDepth": [number],
    "avgRMS": [number],
    "peakLevel": [number],
    "oneSentenceSummary": [string]
  },
  "overallFeedback": [string],
  "audioQualityFeedback": {
    "loudness": [object with issue/suggestion/analogy],
    "distortion": [object],
    ...
  },
  "soundDesignFeedback": {
    "timestampedIssues": [array of objects with timestamp/issue/suggestion],
    "generalSuggestions": [array of strings]
  }
}
```

This structured format ensures consistency and allows the interface to display feedback in organized, scannable sections.

Each module operates through a structured prompt template encoded within the SonicLens backend, designed to maintain consistency and transparency in the generated feedback. The complete prompt is assembled as a multi-part message that separates textual instructions from the audio media, allowing the AI to process both simultaneously.

Scope and Deployment

The current implementation of SonicLens functions as a self-contained web platform. It does not integrate with external learning management systems. The system is designed to manage audio uploads, feedback, and revisions within a unified workspace. This ensures that all critique data remains connected and accessible for reflective review throughout the creative process.

The primary objective of SonicLens is to support the iterative nature of the design studio. While the current paper focuses on the technical architecture and theoretical framework of the system, this section outlines the intended deployment strategy and the planned empirical evaluation. The system is designed to complement, rather than replace, the role of the human instructor by providing consistent, data-grounded feedback during the early stages of the creative process.

Intended Use in the Design Studio

SonicLens is intended for use in foundational sound design courses within broader design and creative media programs. Students will utilize the platform to upload drafts of their soundscapes or narrative audio projects. The system is designed to provide immediate formative feedback, allowing students to refine their work before presenting it in formal peer review or instructor-led critique sessions.

This implementation aims to reduce the “feedback gap” that often occurs in large-enrolment studios where instructors cannot provide individualized technical guidance to every student simultaneously. By interacting with the AI-generated metaphors, students are encouraged to engage in self-reflection. They can then bring specific, technically-informed questions to their human mentors, fostering a deeper and more productive dialogue during physical studio hours.

Evaluation Framework

The evaluation of SonicLens will focus on two primary dimensions: pedagogical effectiveness and technical accuracy. To satisfy the need for empirical validation, a formal study is planned for the upcoming academic semester. This study will utilize a mixed-methods approach to capture both quantitative learning gains and qualitative user experiences. The commencement of this study is contingent upon securing the necessary institutional ethical approval.

The evaluation will be guided by the following research questions:

1. To what extent does the use of visual metaphors improve a student's ability to identify and correct technical audio issues?
2. How does the presence of explainable AI (XAI) feedback influence a student's trust and engagement with technical principles?
3. In what ways does the system facilitate a more collaborative critique environment between students and instructors?

Planned study design

The upcoming study is expected to involve a cohort of at least 30 design students enrolled in a sound design module. This sample size is intended to provide sufficient data to identify patterns in feedback uptake and comprehension. Following ethical clearance, participants will be invited to join the study and will be divided into two groups: a control group receiving traditional technical feedback and an experimental group utilizing the SonicLens platform with metaphorical scaffolding.

Quantitative data will be collected through pre-test and post-test assessments of audio technical knowledge. Students will be asked to identify specific technical issues in provided audio samples both before and after using the system. Furthermore, "feedback uptake" will be measured by analysing the frequency with which students implement AI-suggested revisions in their final project submissions.

Qualitative Metrics and Success Indicators

Qualitative data will be gathered through semi-structured interviews and focus groups. These sessions will explore how students perceive the visual metaphors and whether the analogies helped them overcome the "translation gap" identified in the literature review.

The success of the SonicLens intervention will be measured against the following indicators:

- **Comprehension:** A measurable increase in the accurate use of technical audio terminology in student reflections.
- **Reflective Iteration:** An increase in the number of project revisions performed by students compared to previous cohorts.
- **Metaphor Resonance:** The degree to which students adopt the visual metaphors in their own descriptions of their auditory work.
- **Trust and Agency:** The extent to which students feel confident in accepting or rejecting AI suggestions based on the provided explainable rationale.

By establishing this rigorous evaluation plan, we aim to validate the theoretical and technical contributions presented in this manuscript. The results of this study

will inform future iterations of the metaphor library and the refinement of the prompt-chain logic.

Findings

As a design and system report, the findings presented in this section are framed as the implemented system features, interface design decisions, and pedagogical affordances that emerged from the development of SonicLens. Empirical findings from student use will be reported in a subsequent study, as outlined in the Methodology section. The following subsections present these design findings, organized around the core interface components and their pedagogical rationale.

The core interface of SonicLens is designed to provide formative feedback that is both technically grounded and conceptually accessible. After the analysis pipeline is complete, the system presents the feedback directly onto a waveform view. This layout allows students to synchronize auditory events with specific AI observations. As shown in Figures 4–6, the feedback is structured into three distinct panels that separate the overall summary, technical audio quality metrics, and qualitative sound design suggestions.

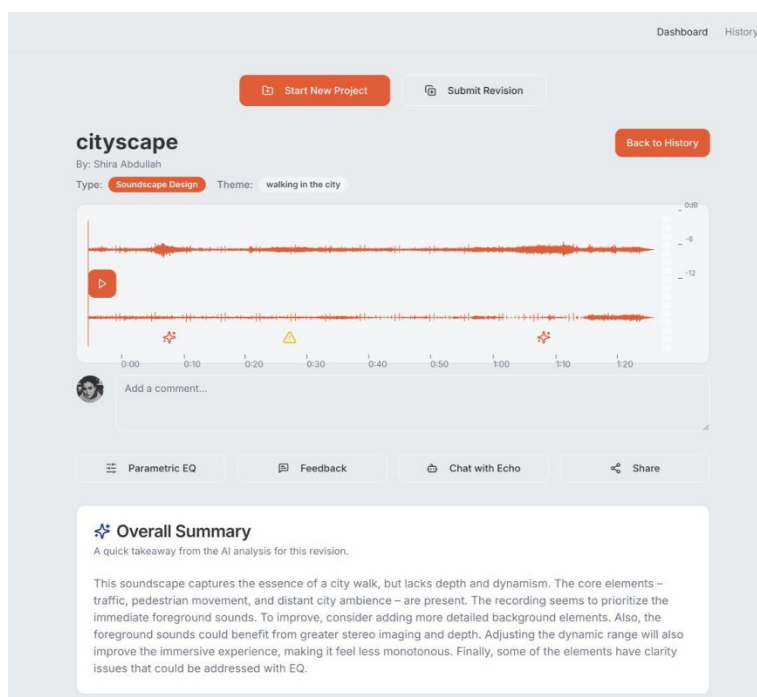


Figure 4. SonicLens Waveform view and overall summary analysis

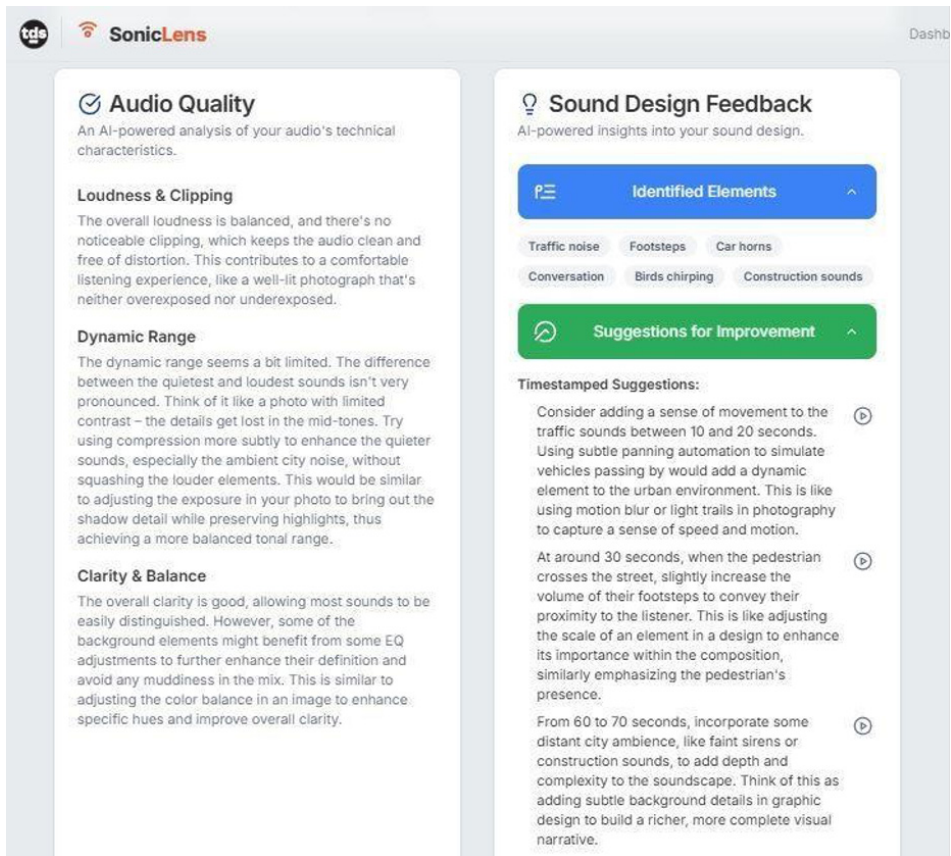


Figure 5. Lower half of the page displaying structured panels separating audio quality issues from sound design suggestions

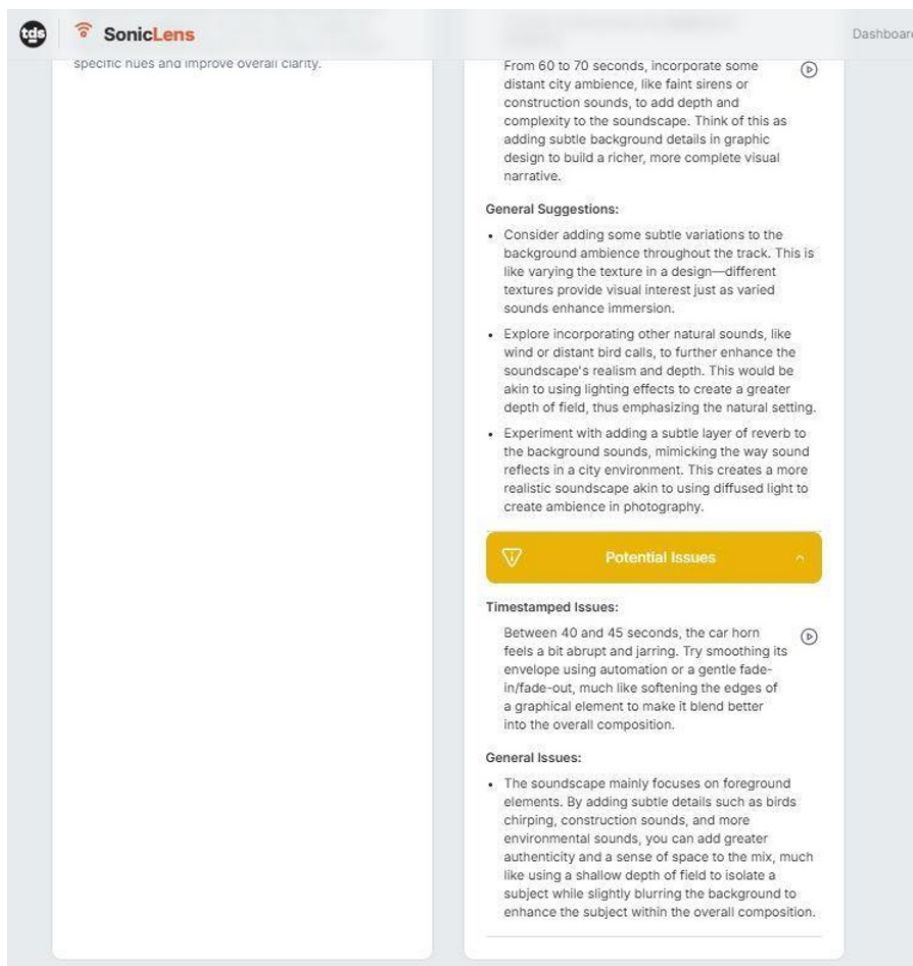


Figure 6. Bottom part of the page showing remaining of the AI suggestions and potential issues

Explainable Interface and XAIxArts

The transparency of the feedback is a central design requirement. SonicLens utilizes the XAIxArts framework to ensure that every suggestion includes a clear rationale. Each AI-generated comment provides a short explanation that identifies the specific segment and data that triggered the feedback (Figure 7). This includes references to measurable trends such as loudness peaks or spectral shifts. By clicking on specific markers in the waveform, students can replay the corresponding audio segment while reading the AI's reasoning in a contextual tooltip.



Figure 7. Waveform interface showing AI suggestions with reasoning tooltip. Each marker indicates a timestamped segment where explainable feedback is anchored to the analyzed audio.

This approach aims to move beyond automated correction toward a form of pedagogical scaffolding. Rather than simply stating that a track is too loud, the system explains why the current levels might lead to distortion, drawing a connection between the raw data and the creative outcome. This level of transparency is intended to build trust and encourage students to engage more deeply with technical principles.

Visual Metaphors for Non-Audio Specialists

As established in the literature review, visual metaphors are utilized to bridge the translation gap for design students. When the system identifies a technical issue, it translates the concept into a visual design analogy. For example, a suggestion regarding dynamic range is presented alongside an analogy to layout contrast. A comment on spectral balance might be compared to colour palette harmony.

These metaphors are designed to appear alongside the technical analysis so that students can interpret sound through their existing visual literacy. By referencing the mappings established in Table 1 (Section 2.2), the system provides a familiar conceptual frame for abstract auditory phenomena. This method is intended to reduce cognitive load and allow students to make more confident decisions during the revision process.

Integration of Human Critique

SonicLens is designed as a human-in-the-loop system where AI feedback exists alongside human expertise. The interface allows instructors and peers to add their own timestamped comments directly onto the same waveform (Figure 8). This creates a shared space for critique where students can compare machine-generated insights with human perspectives. The combination of these sources, displayed together in the feedback panel (Figure 9), helps students understand not only the technical requirements of their project but also the subjective and creative nuances of sound design.

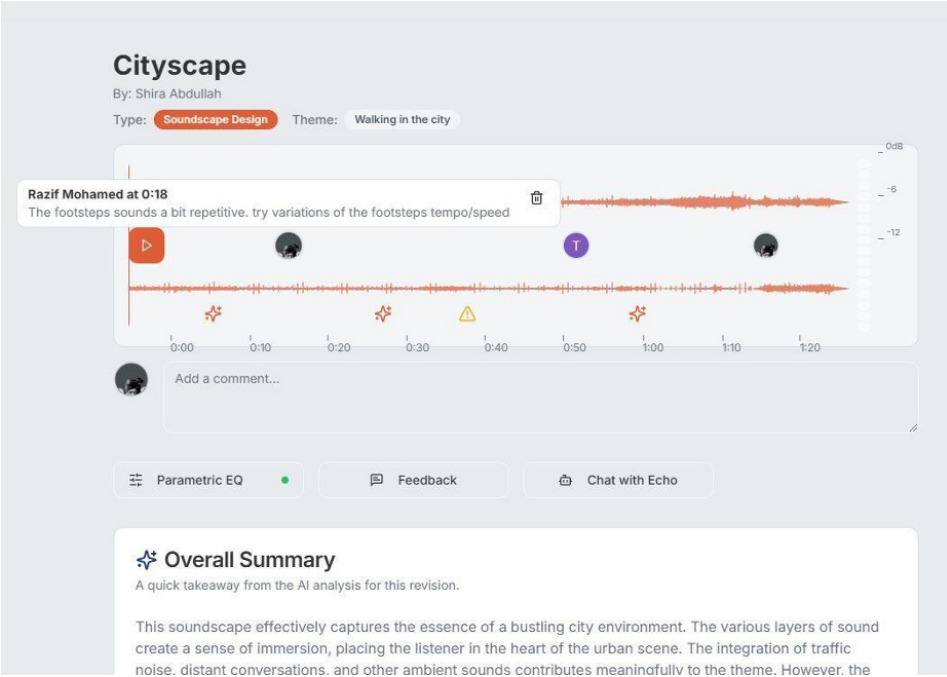


Figure 8. Instructors’ and peers’ feedback on the same waveform interface

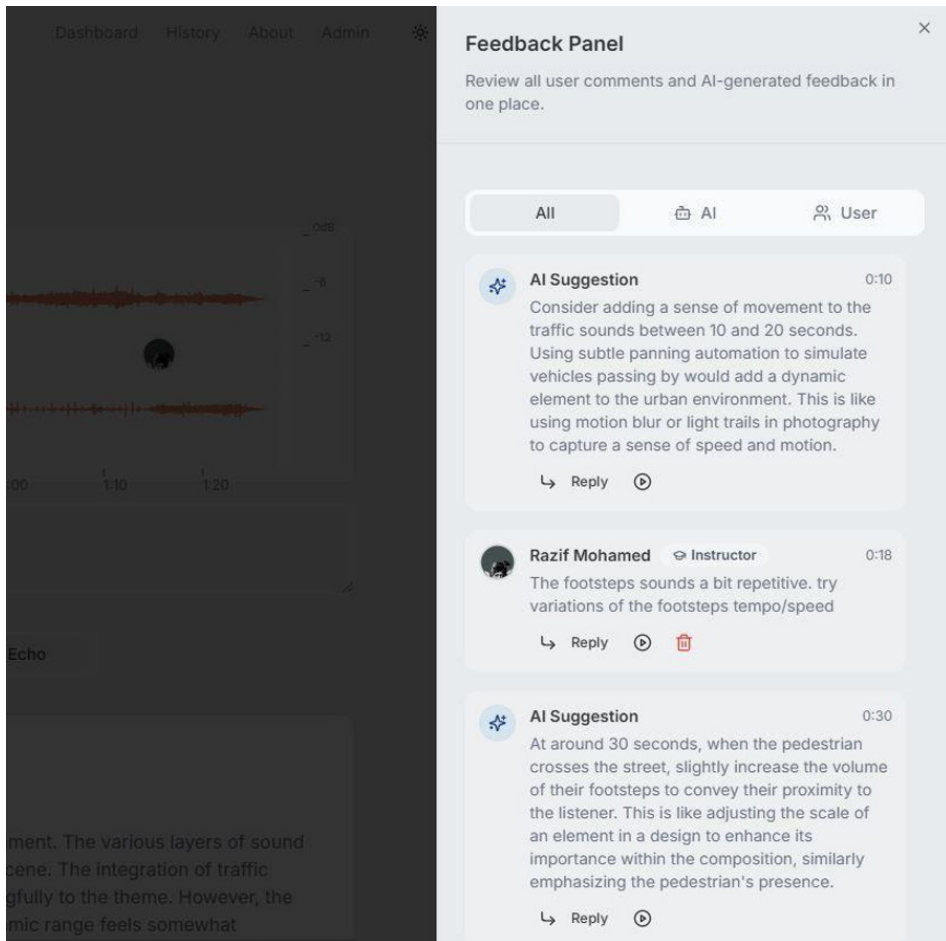


Figure 9. Feedback panel view displaying both AI and instructor comments. The shared thread design encourages comparison, discussion, and collaborative reflection

Collaborative Dialogue and Reflection

The feedback panel encourages students to respond to the suggestions they receive. Students can use this space to reflect on the AI's analysis or to express disagreement with a particular suggestion. These responses are organized into threaded discussions where instructors and peers can join the conversation (Figure 10). This dialogue turns what begins as an automated critique into a social learning experience. By requiring students to justify why they accept or reject a suggestion, the system encourages metacognitive awareness and critical thinking.

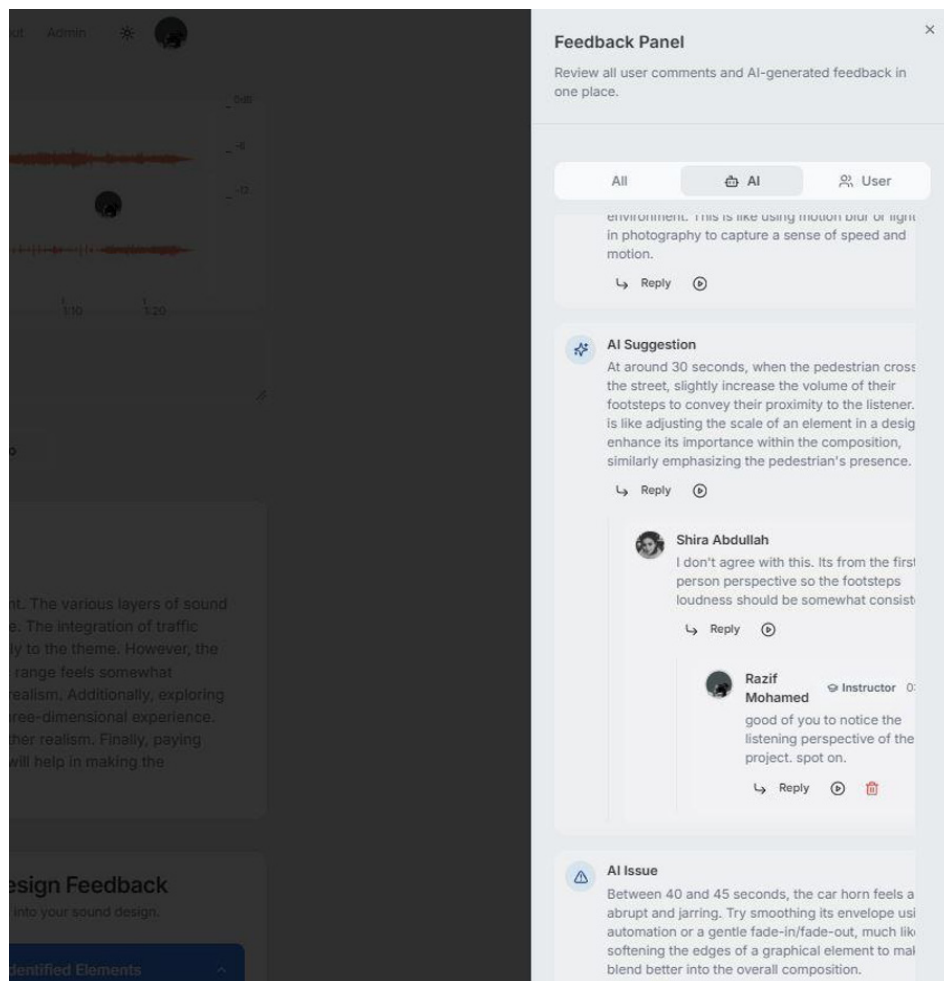


Figure 10. Students and instructor engaging in a discussion

Technical Scaffolding through Parametric Equalization

To further support the learning process, the system includes a built-in Parametric Equalizer. Instructors can use this feature to share specific EQ pre-sets as suggestions for overall audio enhancement. Students can load these pre-sets to make direct comparisons between their original mix and the instructor's suggested version (Figure 11). This hands-on comparison is intended to sharpen critical listening skills and provide a practical understanding of how frequency adjustments influence the overall quality of a soundscape.

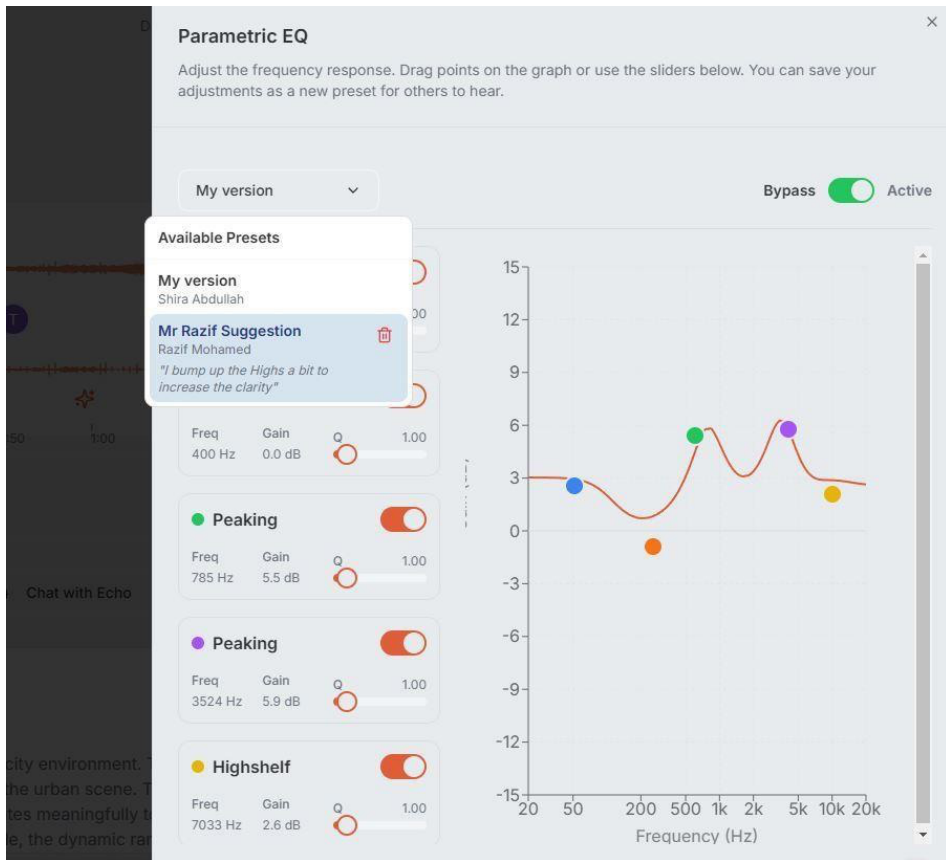


Figure 11. Instructor's Parametric EQ presents suggestions to students

Discussion

The development of SonicLens suggests that artificial intelligence can serve as a conceptual bridge between technical audio parameters and visual design intuition. By translating abstract auditory data into familiar metaphors, the system allows students to engage with their work through a cognitive lens they have already mastered. This approach moves away from traditional visualization tools, which require expert decoding, toward an interpretative model that supports novice learning.

Comparison with Contemporary Systems

SonicLens occupies a distinct position within the emerging landscape of AI-augmented sound design tools. It shares conceptual similarities with the Audio Metaphor (AuMe) system, which also utilizes metaphorical language to support

creative workflows (Liu et al., 2025). However, while AuMe focuses primarily on the generation of soundscapes, SonicLens is specifically designed for formative feedback and critique. By providing timestamped, explainable rationale alongside its metaphors, SonicLens aligns with Lim et al.'s (2023) conception of AI as a “relational artifact” that mediates reflective dialogue rather than delivering authoritative judgment. This framing positions the AI not as an automated judge, but as a technical collaborator that facilitates deeper reflection between the student and their creative intent.

The Mapping Problem and Cultural Variability

A central challenge in the design of the system is the “mapping problem,” which concerns the relationship between auditory features and their visual representations (Roddy & Bridges, 2020). While the metaphors in Table 1 are grounded in cross-modal research, they are not universally perceived. Cultural background, age, and prior training can significantly influence how an individual associates a visual shape or colour with a sound quality (Walker, 1987). The current metaphor library in SonicLens reflects specific Western design conventions that may be less intuitive for students from different cultural contexts.

Recognizing this limitation is essential for ensuring that the system remains an inclusive pedagogical tool. Future iterations of the platform could allow for customizable metaphor libraries that adapt to the specific cultural and disciplinary backgrounds of the user.

Technical and Pedagogical Limitations

There is a risk that students may view AI-generated feedback as an absolute technical truth rather than a creative provocation. An overreliance on metaphorical scaffolding can lead to the oversimplification of the complex physical properties of sound (Pirhonen, 2014). For example, while comparing spectral balance to colour harmony aids initial comprehension, it does not fully capture the mathematical nuances of frequency masking. Furthermore, the system is currently limited by the training data of the underlying language model, which may struggle to evaluate avant-garde or experimental compositions that intentionally violate traditional audio standards. To mitigate these risks, SonicLens is designed as a human-in-the-loop system. The interface prioritizes the presence of instructor and peer comments to ensure that the machine's perspective remains one part of a broader, social critique.

Ethical Considerations and Authorship

The integration of AI into creative assessment raises significant questions regarding student agency and authorship. SonicLens addresses these concerns by adopting the

XAIxArts framework (Bryan-Kinns et al., 2023), which emphasizes transparency and explanation as core principles when AI operates in creative contexts. By providing the technical rationale behind every suggestion, the system encourages students to critically evaluate the feedback. This process is intended to foster “metacognitive agency,” where students learn to accept or reject AI suggestions based on their own design goals rather than blind compliance.

Conclusion

SonicLens represents a pedagogical innovation that seeks to bridge the translation gap in sound design education. By merging automated technical analysis with explainable visual metaphors, the system provides a structured environment for iterative reflection and critique. The modular prompt-chain and tool-calling architecture ensure that the feedback remains technically grounded, while the cross-modal mappings make technical concepts accessible to design students.

As a design and system report, this paper establishes the theoretical and technical foundation for the SonicLens platform. While the system shows promise as a tool for scaffolding technical comprehension, formal empirical validation remains the critical next step. The planned study, involving a cohort of at least 30 students, will provide the necessary data to evaluate learning gains and student engagement. Ultimately, SonicLens aims to extend the reach of the design studio, providing students with the continuous, data-informed guidance required to master the invisible complexities of sound.

Recommendations for Future Research

Future research on SonicLens should focus on three interconnected directions. First, empirical validation through the planned evaluation study will be essential to assess the system’s pedagogical effectiveness across diverse student cohorts, particularly with regard to comprehension gains, metaphor resonance, and reflective iteration. Second, the system’s metaphor library should be expanded and culturally adapted to address the variability in cross-modal interpretation across different student populations, including those from non-Western design traditions. Customizable metaphor sets and instructor-curated analogies would allow the platform to accommodate diverse pedagogical contexts. Third, longitudinal studies are recommended to examine how sustained use of SonicLens influences students’ long-term development of audio literacy and design vocabulary. Future iterations should also explore integration with learning management systems, multi-language support for Southeast Asian educational contexts, and the inclusion of additional creative modalities such as video and interactive media. Finally, further investigation into the ethical dimensions of AI-mediated critique, including authorship attribution and student agency in accepting

or rejecting AI feedback, will be critical to ensuring that systems like SonicLens remain transparent, trustworthy, and pedagogically sound.

Open Access: This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0) which permits any use, distribution and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Bryan-Kinns, N., Ford, C., Chamberlain, A., Benford, S., Kennedy, H., Li, Z., Wu, Q., Xia, G. G., & Rezwana, J. (2023). Explainable AI for the arts: XAIxArts. In *C&C '23: Proceedings of the 15th Conference on Creativity and Cognition* (pp. 1–7). Association for Computing Machinery. <https://doi.org/10.1145/3591196.3593517>
- Crowther, P. (2013). Understanding the signature pedagogy of the design studio and the opportunities for its technological enhancement. *Journal of Learning Design*, 6(3), 18–28. <https://doi.org/10.5204/jld.v6i3.155>
- Duan, P., Chen, C.-Y., Li, G., Hartmann, B., & Li, Y. (2024). UICrit: Enhancing automated design evaluation with a UI critique dataset. In L. Yao, M. Goel, A. Ion & P. Lopes (eds), *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)* (pp. 1–17). Association for Computing Machinery. <https://doi.org/10.1145/3654777.3676381>
- Elmosnino, S. (2022). A review of literature in critical listening education. *Journal of the Audio Engineering Society*, 70(5), 328–339. <https://doi.org/10.17743/jaes.2022.0004>
- Fahlenbrach, K. (2008). Emotions in sound: Audiovisual metaphors in the sound design of narrative films. *Projections*, 2(2), 15–32. <https://doi.org/10.3167/proj.2008.020206>
- Giannakis, K. (2006). A comparative evaluation of auditory–visual mappings for sound visualisation. *Organised Sound*, 11(3), 297–307. <https://doi.org/10.1017/S1355771806001531>
- Gray, C. M., & Howard, C. D. (2015). Critical thinking, critique, and signature pedagogies in design education. *Dialectic*, 1(1), 1–18. <https://doi.org/10.22492/ijeds.3.1.01>
- Hicks, K., Rogers, K., Gerling, K., & Nacke, L. (2024). Juicy audio: Audio designers' conceptualization of the term in video games. In J. Nichols (ed), *Proceedings of the ACM on Human-Computer Interaction (Vol 8)* (pp. 1–28). Association for Computing Machinery. <https://doi.org/10.1145/3677084>
- Joo, W. (2022). Graphic-to-sound sonification for visual and auditory communication design. In *Proceedings of the 17th International Audio Mostly Conference (AM '22)* (pp. 1–6). Association for Computing Machinery. <https://doi.org/10.1145/3561212.3561214>
- Lemaitre, G., Scurto, H., Françoise, J., Bevilacqua, F., Houix, O., & Susini, P. (2017). Rising tones and rustling noises: Metaphors in gestural depictions of sounds. *PLOS ONE*, 12(7), e0181786. <https://doi.org/10.1371/journal.pone.0181786>
- Lim, J., Leinonen, T., Lipponen, L., Lee, H., DeVita, J., & Murray, D. (2023). Artificial intelligence as relational artifacts in creative learning. *Digital Creativity*, 34(3), 192–210. <https://doi.org/10.1080/14626268.2023.2236595>

- Liu, G., Lee, K. J., Thorogood, M., Anderson, C., & Philippe, P. (2025, September 10). AI-assisted sound design with audio metaphor (AuMe): An evaluation with novice sound designers. *The AI Music Creativity Conference (AIMC)*, Brussels, Belgium. <https://doi.org/10.5281/zenodo.16946064>
- Magnusson, T. (2017). Musical organics: A heterarchical approach to digital organology. *Journal of New Music Research*, 46(3), 286–303. <https://doi.org/10.1080/09298215.2017.1353636>
- McCarthy, C. (2011). Redesigning the design crit (Research report). *Ako Aotearoa National Centre for Tertiary Teaching Excellence*. <https://ako.ac.nz/knowledge-centre/redesigning-the-design-crit>
- Memarian, B., & Doleck, T. (2024). A review of assessment for learning with artificial intelligence. *Computers in Human Behavior: Artificial Humans*, 2(1), 100040. <https://doi.org/10.1016/j.chbah.2023.100040>
- Pirhonen, A. (2014). Metaphor as a focal concept in sound design. In *AM '14: Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound* (pp. 1–6). Association for Computing Machinery. <https://doi.org/10.1145/2636879.2636894>
- Roddy, S., & Bridges, B. (2020). Mapping for meaning: The embodied sonification listening model and its implications for the mapping problem in sonic information design. *Journal on Multimodal User Interfaces*, 14, 143–151. <https://doi.org/10.1007/s12193-020-00318-y>
- Saritepeci, M., & Yildiz Durak, H. (2024). Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study. *Education and Information Technologies*, 29(18), 25175–25209. <https://doi.org/10.1007/s10639-024-12829-2>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, 134(3), 52–59. <https://doi.org/10.1162/0011526054622015>
- Somers, E. (1998). A pedagogy of creative thinking based on sonification of visual structures and visualization of aural structures. In *Proceedings of the International Conference on Auditory Display (ICAD '98)*. <https://doi.org/10.14236/ewic/AD1998.27>
- Walker, R. (1987). The effects of culture, environment, age, and musical training on choices of visual metaphors for sound. *Perception & Psychophysics*, 42(5), 491–502. <https://doi.org/10.3758/BF03203067>
- Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound–colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, 42(2), 264–280. [https://doi.org/10.1016/S0010-9452\(08\)70354-X](https://doi.org/10.1016/S0010-9452(08)70354-X)
- Yong, T., & Lee, Y. L. (2026). Fostering peer interaction through a buddy system: Action research in a kindergarten classroom. *Asia-Pacific Journal of Futures in Education and Society*, 5(1), 75–98. <https://doi.org/10.58946/apjfes-5.1.P5>
- Zhu, Y., & Krish, P. (2026). Transforming music lessons with Piascore and EarMaster at a tertiary institution in China. *Asia-Pacific Journal of Futures in Education and Society*, 5(1), 21–40. <https://doi.org/10.58946/apjfes-5.1.P2>