

## Audio-Responsive Architectural Visualization: Linking Musical Features with Real-Time Digital Rendering”

Shujing Tang<sup>1</sup>

### Abstract

The tendency in the visualization of architecture is better observed in the transformation of the architectural visualization as inert image to the multisensory and active digital mediation. Audio-responsive visualization is one of these new directions, providing a contentious conceptual base on which to reconsider the process of corporate translation of spatial perception, affect and environmental awareness into real-time architectural imagery. In this study, the researcher presents the concept of a hybrid conceptual-technical model that can be used to connect musical feature extraction and adaptive rendering rules so that the visual outputs can evolve continuously depending on the sonic features. At the crossroads of computational design, architectural representation, interactive media, and phenomenology, the paper outlines the theoretical basis as well as the system process by which sound-based visualization can be accomplished. The conceptual part investigates the role of rhythm, timbre, frequency, and amplitude in the formulation of spatial interpretation of the musical qualities that could be brought forth as generative stimulants in the architectural imaging logics. The technical part presents a prototype workflow described with short time fourier transforms (STFT), spectral centroid tracking, rhythmic onset detection and energy mapping to affect color dynamics, geometry deformation, particle behavior and environmental lighting. The hybrid paradigm is then understood through the representational agency, digital materiality and human perceptual agency. These findings imply that audio-responsive visualization extends the architectural visualization beyond traditional photorealism and still composition to produce images that are immersive, time-based and full of affect, and appeal to modern representational cultures. The research paper then provides a conclusion on the implications of the study to real-time rendering practice, design pedagogy, public installations, and future interdisciplinary research at the convergence of architecture, sound studies, and computational media.

**Keywords:** *Audio-Responsive Visualization; Architectural Representation; Digital Rendering; Sound-Visual Mapping; Real-Time Systems; Computational Design; Interactive Architecture; Multisensory Environments; Digital Aesthetics; Spatial Perception.*

### Introduction

Traditionally, architectural visualization has been occupied by immobile drawings, unvarying frame images and the curated simulation of photographs that show space as something that can be consumed, evaluated or imagined. These representational traditions, despite their refinement, are attached to a visual regime that keeps architecture apart, in terms of time, sense, and feeling, of how it is really lived. The modern culture of media, though, is more and more inclined to require the types of representation that would be able to recognize the fluidity, dynamism and multisensory immersion. With the explosion of digital technologies to extend the limits of visualization, architecture is facing the challenge of utilizing the time, movement and sound not as the peripheral decorations but as significant elements of spatial meaning.

One of the aspects where experimentation is made in such a changing environment is audio-responsive architectural visualization. It implies that musical forms (rhythm, frequency patterns, amplitude envelopes and textual qualities) could determine the generative tendency of real-time digital renderings. Instead of considering sound as decorative elements in an atmosphere displayed upon a visual pattern, this type of treatment is placing sound as a dynamic parameter, which co-writes the

---

<sup>1</sup> Senzoku Gakuen College of Music, Composition Major of the Graduate School of Music, Kanagawa, 213-8580, Japan, Email: Tshujing6666@gmail.com

image. Pictorial releases are performative, adaptive and positional within temporal patterns of sound-based input. An architectural rendering does not freeze any moment, but it is an ever-changing phenomenon (Gancz et al., 2025).

The study represents a composite conceptual-technological study in keeping with the requirements of *Architecture Image Studies*, a journal that has been prominent in permitting interdisciplinary interactions with architectural media, architectural representation, and technological culture. The conceptual direction considers the impacts of sound on spatial perception, the possibilities of abstracting musical elements into representational stimuli, and how multisensory awareness can be reformulated to fall under the practices of visualization. The technical direction presents a system workflow, based on the modern tools of audio analysis and rendering, illustrating the possibility of associating the extraction of musical features with dynamic image adjustment. Collectively, they form a complete system of audio-reactive visualization as a valid representational approach.

### **Architecture, Sound and Perception.**

A sound to spatial experience relationship is not a new concept in architectural theory. Sound determines the scale, the density of materials, closeness, and enclosure. Spatial dimensions are manifested in the reverberation patterns. Beats create an accent of acceleration of sensation of movement or deceleration of sensation movement. The timbre adds to the character of the atmosphere. But in the computationally significant sense, sound is uncommonly part and parcel of architectural visualization despite this profound reciprocity. The visual is independent, disconnected with the sound.

Visualization, which is responsive to audio, disputes this division. It claims that when architecture is inhabited as a multisensual experience, then images in which architecture is visualized should develop as multisensory too. Sound is particularly a powerful medium in that it is inherently time-based, flowing and emotionally communicative. Its form can give time-logic to visual expression, to make images read out rhythmically, fluctuate energetically, and breath with audible sonic forms.

### **Digital Rendering Digital Rendering as a Dynamic Mediation**

Conventionally, the rendering engines recreate light, material and geometry to generate images that can be close to the perceived world. Rendering in real-time Rendering is dynamic in real time, that is, in gaming engines, interactive installations, and VR, wherein it is also reacted to by the movement of the user or the environment. Increasing the responsiveness to sound builds on the computational infrastructures that are available but changes the representational result. (Huang et al., 2025a)

Architectural imagery is algorithmically reactive when musical characteristics are considered as military signals of visual characteristics. Edges may be sharpened or random noise of particles may be introduced by the high-frequency peaks. The slow geometric deformations can be induced by low-frequency bass. Amplitude spikes can occur suddenly to enhance the luminance or to trigger the effects of spatial ripple. Constant harmonious sounds can alter color schemes into gradations. Instead of depending on random visual effects, the system renders logic of sound structure as the interpretable and rule-based system (Popp & Murphy, 2022).

### **Moving towards a Multisensory Theory of Architectural Images.**

The presence of sound as a generative tool, forces one to redefine what an architectural image is. It is not a mere repository of information anymore, but a performative system, whose properties can be formed in the interactions. In such a framework:

- Images are temporal: they evolve over the time of musical stimulus.
- The pictures are relational: they are based on the correspondence between the sonic and visual features.
- The images are performative: they are not the representation of space but they act in a spatial way.
- Images are contingent: they never achieve full stability, and rebel against the crystalline perfection of photorealism.

These changes are in line with the wider changes in architectural representation; no longer to be based on fixed idealization, but to manifest as expressive media ecologies in which images, tools, and processes develop together.

The conceptual-technical framework hybrid chosen in the current study is in line with the current expectations in research in the study of architectural media. A strictly conceptual essay would run the risk of being abstracted without being operational, and a strictly technical engineering paper would be further out of the representational purview of AIS. The hybrid model enables:

- Perceptual, representational, digital aesthetic, and multisensory theoretical foundation.
- Technical feasibility by having well defined system to extract and visualize real time features.
- Critical discourse analysis on the way such a system transforms the agency of the architectural imagery.
- Relation to practice by pointing out implications to design learning, interactive installations, public displays, and computation processes.
- By uniting these dimensions, the study provides a unique representational approach, connecting the practice of artistic, architectural, and digital media.

The main goal of this study is to design and test a hybrid conceptual-technical model that can connect musical characteristics and adaptive architectural rendering. The research has four objectives that are interlinked:

- Theoretical sound as a representational operative in architectural visualization.
- To determine which musical characteristics (e.g. rhythm, spectral density, amplitude dynamics) would be most appropriate to real-time mapping.
- To build a system workflow that is able to extract features and implement rule-based rendering transformations.

To derive the meaning of this approach to the architectural image culture.

## **Literature Review**

Audio-responsive architectural visualization exploration can be identified within a number of overlapping fields of knowledge: multisensory perception in architecture, audio-visual mapping in digital media, computational design and interactive systems, and the development of architectural representation theories. This part brings together these areas in order to place the suggested hybrid model into a logical intellectual and technological tradition.

The two senses that interact in this context are tactile and visual senses. 2.1 Multisensory Perception and Architectural Experience.

There is no vision of architecture. There has been significant phenomenological and cognitive research highlighting that spatial perception is created in the framework of a multisensory continuum in which sound, light, materiality, and movement come to meet. Classical phenomenologists including Merleau-Ponty anticipate in advance the foregrounding of perception as embodied and dynamic, thus implying that spatial significance is created through the interaction of the senses and modalities of perception and not through visual signals. Modern theorists of architecture build on this position and claim that sound significantly influences the atmosphere in space and leads to environmental legibility (Kalidas, 2023).

Research on acoustic ecology shows that auditory stimuli can be used to yield information about enclosure, scale, surface hardness and distance in many cases more quickly and intuitively than visual stimuli. Volumetric characteristics can be known by reverberation times, diffusion patterns and frequency absorption. Depth and mass are conveyed by low frequency waves whereas texture or openness is indicated by a high frequency response. The features are so necessary in comprehending space, but the visualization of architecture to this day still favors photorealistic imagery but the visualization of a sonic perception is peripheral. The lack of this capacity highlights a representational lapse which the audio responsive rendering could offer by structurally including sound into the mediation of architecture.

### **Sound as Generative in the Space and Media Arts.**

Sound has also been widely used as a generative tool in visual arts as well as performance and digital installations, outside architecture. The experimental work on synesthesia-inspired composition attempted to establish the mapping of musical qualities to color or shape. Mid-century kinetic and

electronic art artists came up with mechanical and light systems, which were connected to sound waves. Real-time audio analysis has become a regular part of media installations, projection mapping, live coding, and performances in more recent years. These works usually result in two strategies:(Katan, 2012)

Direct signal mapping, in which amplitude or frequency is the visual controller;

Interpretive mapping, in which sonic features are coded into generative behavior, usually symbolic.

Nevertheless, the description of these explorations rarely touches upon the architectural representation. They center on experiential art as opposed to mediation of designed space. In architectural visualization, the task is not only to make the visual effects of sound-reactive desirable but to make sure that sonic influence is not in vain within the logic of space, it is necessary to modulate light, geometry or environmentality so that it enhances architectural readability, but not kills it.

### **Digital Rendering and Real-Time Visualization**

The emergence of game engines and interactive visualization systems has changed the nature of architectural imaging to be dynamic. Real-time photorealistic rendering using tools like Unreal Engine, Unity, the Eevee renderer of Blender, and WebGL pipelines can deal with user navigation, environmental parameters, or scripted behaviors. Procedural geometry, shader-based effects, and parametric rules are supported by these systems, and it makes them perfect in incorporating audio-responsive inputs.(Morana, 2023)

The use of real-time rendering research focuses on continuity, loops, and adaptive algorithms. Visual variations happen in sets of frames 30-120 fps, and permit subtle or extreme changes timed to the external stimuli. Although the environments are becoming more popular with architects as a way to walk through and have VR experiences, sonic input is hardly ever utilized outside of background audio. This literature reveals a conceptual gap: Multisensory responsiveness can technically be achieved through real-time rendering and is not represented in architectural visualization as a representational paradigm. This lack provides a possibility of study that integrates the logic of rendering with musical analysis to develop active architectural images.

### **Extraction of Audio Features and Computational Mapping.**

Audio-responsive visualization relies on the powerful approaches to the analysis of musical cues. Modern audio studies provide an advanced complex of algorithms that can be used to determine structural and expressive attributes of sound:

- Amplitude envelope- shows dynamics of loudness which are applicable in scaling of visual intensity.
- Spectral centroid gives an insight of how bright a color appears which can be used to adjust color temperature or shader texture.
- STFT has been created to deconstruct sound to form time-frequency elements, allowing multi-band visual effects.
- Rhythmic Onset Detection This detects beat structures and tempo, which is suitable to initiate periodic visual events.
- Spectral flux records variations in harmonic content with time, which are suitable at the time of changing textures or particle behaviors.
- Mass, resonance and depth are associated with low-frequency energy, which implies that the geometric deformation can be mapped.(Vargas, 2025)

The difficulty lies in the process of converting these features into non-trivial and non-arbitrary visual rules. Though the literature of computing and media arts offers many examples of sonification and visualization structures, architectural visualization requires mappings that do not obstruct spatial legibility, atmospheric coherence, and material logic. This involves coming up with rule systems that react to sound without violating architectural representation as cultural and communicative activity.

### **Synesthetic and Cross-modal Interpretation.**

The visualization by audio-responsiveness is further theorized in cross-modal research. It is proven by neuroscientific research that humans have an automatic habit of relating visual qualities to acoustic qualities: bright colours to bright high frequencies, smooth gradients to low frequencies, dense textures to noisy spectra. Although not universal, these correspondences have high statistical consistency. The findings of this kind confirm the habit of the mapping of musical features on the visual ones, not as arbitrary artistic decisions but as perceptually intelligible transformations. (Raposo & Soares, 2025)

The architectural theoreticians point out that it is cross-modality that enriches spatial imagination. Images are gesturing at lived experience when sound-responsive dynamics are encoded in them, but not represented on a static basis. This echoes with the atmospheres, affect and immersion theories in modern spatial studies. Literature indicates that multisensory correspondence helps to increase engagement, embodied interpretation, and the possibility of viewers to intuitively obtain spatial qualities.

### **Architectural Representation in the Digital Media Age.**

The development of architectural representation has been characterized by the replacement of the hand drawing by the digital production, the perspective by the animation, the rendering by the simulation, and currently, by the responsive and interactive environment. According to scholars, this change can be described as the transition from representation to mediation; architecture no longer needs to be drawn but acted out using digital tools. Images are interfaces, protocols, and dynamic fields as opposed to documents.

In this respect, audio-responsive visualization will be part of what scholars refer to as the expanded image. It brings about a notion of temporality, contingency, and multisensory correspondence to representational practice. The architectural image is a never-ending process that is directed by algorithmic behavior. This echoes in the modern discourse of digital materiality, where representations are interpreted as dynamic, procedural and coded with computational logic.

Some of the authors believe that digital representations of architecture should be read not only for what they describe but how they work - how they embrace time, interaction, and environmental parameters. Audio-responsive visualization is in line with this turn of the operation. It enhances architectural imaging by including real-time stimuli and placing the digital representation in a wider ecology of basal ecological conditions of time and senses. (Yu, 2025)

### **Literature Blanks and Justification of the Research.**

Although there are extensive collections of sound studies, media arts, interactive design, and digital rendering, there are several gaps:

Absence of interaction with architectural visualization.

The vast majority of sound-reactive systems are artistic or performance-based. Spatial representativeness and architectural readability is not something they talk about too often.

- Lack of conceptual models between sound and architectural image.
- No literature defines audio-responsive visualization as a representational theory based on architectural perception.
- Scarcity of description of rule-based sonic mappings of spatial imagery.
- Media arts literature usually has an aesthetic experimentation as opposed to a definite mapping logic fitting in architectural representations.
- Inadequate hybrid approaches.
- Not many studies involve a combination of conceptual analysis and functional system workflow applicable to the representational practice.
- Failure to use multisensory intention in architectural teaching and practice.
- Multi-sensory design interfaces have not been introduced into architecture, even though rendering tools are now capable of being responsive in real-time. (Erdmann et al., 2025)

The existing study fills in these gaps with the proposed conceptual-technical model with a solid theoretical basis, research on perception, computational mapping, and digital representational logic. It

cross-fertilizes disciplines and provides a methodological means of developing audio-responsive architectural images.

### **Overview of the Conceptual Foundations.**

Literature synthesis determines some major principles:

- Sound is a potent tool to organize the time and atmospheric and affective space situations.
- The technologies of audio analysis are well-developed to obtain rich information that can be utilized in the manipulation of visual changes.
- Modern rendering systems offer real time structures that can map sonic behaviour into visual behaviour.
- The new trends of architecture representation are moving towards dynamic, multisensory, and performative representation.

Theory, perception, and computation need to be combined using a hybrid approach to come up with a seamless visualization technique.(Lima et al., 2021)

Such premises justify the creation of a hybrid audio-reactive visualization model in architecture, which places the study on a continuum of current arguments on digital materiality, multisensory forms of representation, and computational design culture.

### **Methodology**

The conceptual-technical logic pattern of the methodology used in the present study is a hybrid one and is representative of modern trends in the research of architectural media. Instead of considering the analysis of sound and visualization as two distinct fields, this methodology incorporates the perceptual theory, cross-modal reasoning, principles of computational mapping, and real-time rendering strategies into one continuous workflow. It is hoped that the goal will be to develop a theoretically and technologically viable, representationally coherent process so that the musical properties can produce significant architectural visual changes. The methodology is thus carried out in an order of conceptual modelling, the creation of a computational pipeline, and the interpretative assessment of the representational outputs of the prototype.

#### **The Conceptual Modelling Framework is Presented**

The conceptual modelling framework defines the theoretical rationale that provides the explanation of why it is reasonable to connect musical properties to architectural visual changes. The principal principle of this scheme is the concept of temporal correspondence, according to which sound is discussed as a time-related phenomenon, the structure of which should be reproduced in the process of visualization. Since music is constantly changing in its traits, the architectural image also needs to act dynamically as opposed to a fixed composition. This time synchronization makes the visual output act as an event that can coexist with the sound stream.(FUKAYAMA, 2020)

Perceptual coherence is another necessary element in the conceptual model. Studies of cross-modal perception have uniformly shown that certain visual properties are intuitively associated with certain sonic ones, e.g., bright colors correspond to high frequencies, or quantal tonalities are perceived as heavier spatially. The perceptual biases provide a conceptual foundation of non- arbitrary mapping between sound properties and visual variables. The methodology thus aims to build mappings that are significant and perceivable so that the visualization reacts to audio input in such a way that such reactions are receptive to human multisensory interpretation.(Tiffon et al., 2022)

Representational legibility is the third component of the conceptual framework that stipulates that the imagery of architecture should not lose spatial coherence in the process of being transformed in audio-driven fashion. In contrast to projects involving media-art, architectural visualization needs to retain recognizable spatial indicators, a logic of materials and safeguard form definition. Based on this, the conceptual model would inform the system to favor atmospheric, lighting, material, or manipulated geometries instead of distortions that would jeopardize architectural knowledge.

#### **System Architecture Overview.**

The system architecture is designed in the form of the multi-layered pipeline connecting the real-time rendering output with the audio input. This structure consists of an input layer to capture and

prepare audio information, an analysis layer to extract informative musical characteristics and an output layer to interpret the extracted characteristics in a real-time visualization setup. The organization of this pipeline is more of modern computational visualization methods, where streams of data are subjected to sequential processing and are also subject to temporal variation. The architecture will be built to be continuously running, which will allow the visualization engine to frame by frame to receive the changing sonic input.

### **Audio Capture and Pre-Processing.**

The methodology starts with audio capture and pre-processing whereby it is ensured that the raw sound signal can be analyzed in real time with computational stability. Regardless of whether it is an offboard recording of a live microphone or a pre-recorded track, the audio is again sampled to a standard rate of 44.1 kHz and normalized to create a standard amplitude range. The methodology uses windowing functions including Hann window or Hamming window to provide the ability to respond to the temporal domain e.g. Hann window or Hamming window to smooth the angularities of cut up audio frames and to limit spectral edge cases. These frames, which normally take twenty to forty milliseconds, enable the system to do the computations of feature extraction quickly enough to support the visual updates in video frame rates. The pre-processing stage converts an unstructured audio signal with the use of segmentation and normalization into a temporally ordered dataset which can be analyzed. (Olowe et al., 2017)

### **Feature Extraction**

The analytical core of the methodology is featuring extraction because the properties found during this step predetermine the visual behaviours that are to occur. The system identifies various musical attributes using Short-Time Fourier Transform (STFT) and complementary detection algorithms which are both computationally and perceptually strong. The amplitude envelope offers a temporal representation of loudness change, and is therefore appropriate to control the intensity of lighting, exposure or material reflectivity of the visualization. Shifts in color temperature or highlight behaviour in architectural surfaces can be directed by the spectral centroid, a description of the brightness of the sound in terms of frequencies distribution.

The system also breaks the audio down into low-, mid-, and high-frequency ratios, and each of these bands has its own potential generative possibilities. Low frequency material can be exploited to create slow geometric vibrations or to modulate the volume of the atmosphere, mid-range frequencies affect the ambience of an environment, and high frequency material is used to create sharp visual details or accent of the particle. Rhythmic onset detection detects beats and impulses that may cause discrete visual events and spectral flux detects change in harmonic content allowing fluid movement of textures or materials according to musical transitions. These aspects together offer a multidimensional data set with the help of which the visualization can work in response and fluently. (Fourney & Fels, 2009)

### **Strategy and Rule Definition Mapping.**

The mapping strategy defines the way the numeric audio characteristics are converted into the visual architectural behaviours. A rule-based system is taken in this study due to its interpretability, clarity, and compatibility with architectural representation. Scalar mappings provide the opportunity to directly scale visual parameters using continuous values, e.g. amplitude or spectral centroid. An example is that with an increase of amplitude, the world lighting can be brightened, or the glossiness of materials can be changed, so that the visualization can display the dynamic outline of the music.

Simultaneously, threshold-based rules can be discrete sonic events like beats or transient spikes that can cause a particular visual transformation. These may be periodic pulses of emissive materials, abrupt exposure changes, or instantaneous spatial vibrations that travel through a facade. A methodology represented as composite rules also combines multiple features to form more complex behaviours such as geometric tessellations being activated when the brightness and loudness are above certain thresholds. The mapping strategy provides an organization yet expressive relationship between sound input and architectural image through sets of rules of varying complexity (Huang et al., 2025b)

### **Rendering Framework**

The rendering structure puts the methodology in an environment of a real-time visualization which is capable of responding in real-time to the incoming audio information. The necessary infrastructure

containing shader editors, dynamic lighting systems, particle engines, and geometry manipulation software is provided by such platforms as Unreal Engine or Unity. Here, the procedural logics represented by the mapping strategy can be transformed into shader functions, vertex movement schemes, material shift and lighting variation.(Robyn Taylor, 2006)

The rendering engine operates with at least sixty frames per second so that correspondence of sound and image is low. It is even possible to manipulate the audio signal in order to make subtle adjustments to the camera system such as making small changes on the exposure, depth of field or motion and all of this adds to the immersive aspect of the visualization. It is the framework of rendering that is, therefore, the point at which the conceptual and computational aspects of the methodology meet, creating an uninterrupted flow of imagery conditioned by sonic dynamics.

### Prototype Workflow Summary

Combining all the above elements, the prototype workflow (Figure 1) starts by taking in audio and pre-processing it, breaking it down, and making it readable to perform analysis. The extraction of features is a continuous process that generates a stream of data that reflects the changing structure of the music. Those features are mapped using rules and sent to the visualization engine, which also updates the architectural parameters in real time. The system consequently creates a living architectural image, the behaviour of which is co-authored by the incoming sound. It is not linear but cyclical and each rendered frame reacts to the latest audio characteristics and preconditions further changes.(Taylor, 2025)

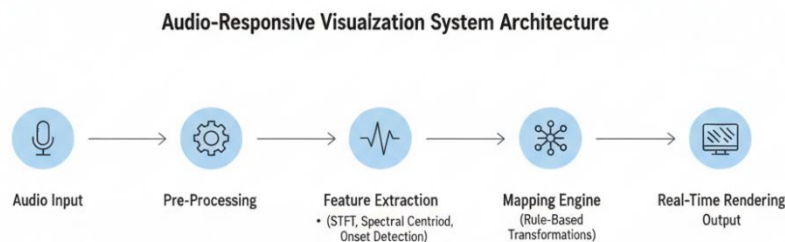


Figure 1. Prototype Workflow Summary

### Evaluation Approach

The critique of the methodology is on representational coherence instead of engineering efficiency. The visualization is studied in terms of the intelligibility of the sound-image relationships, on the one hand, along with the architectural legibility, on the other hand, as well as the depth of the atmospheric nature of the audio-driven changes. The evaluation is based on whether the mappings create behaviorally significant perceptions, whether the system is spatially articulate, and whether the dynamic image is effective in expressing architectural properties instead of fragmenting into some aesthetic visualisation processes. The interpretation methodology is evaluated based on its potential to develop

architectural representation into a multisensory, temporally adaptive medium, through interpretive analysis.

**Limitation**

Although the methodology exhibits great potential, it is recognized that there are a number of limitations. The model gives greater emphasis on musical responsiveness than on architectural acoustics and thus does not model the spatial reverberation or interactions of materials. Real-time performance will always be dependent on the hardware capacity, and certain effects that are computationally intensive might be device-constrained. Also, the mapping rules should be calibrated appropriately since various musical genres generate feature profiles that are too dramatic in nature or not dramatic enough in visual behaviour. With these limitations in place, the methodology offers a solid base on which sound-responsive behaviour can be introduced into architectural visualization and also has a significant input in the representational discourse today.

**Results**

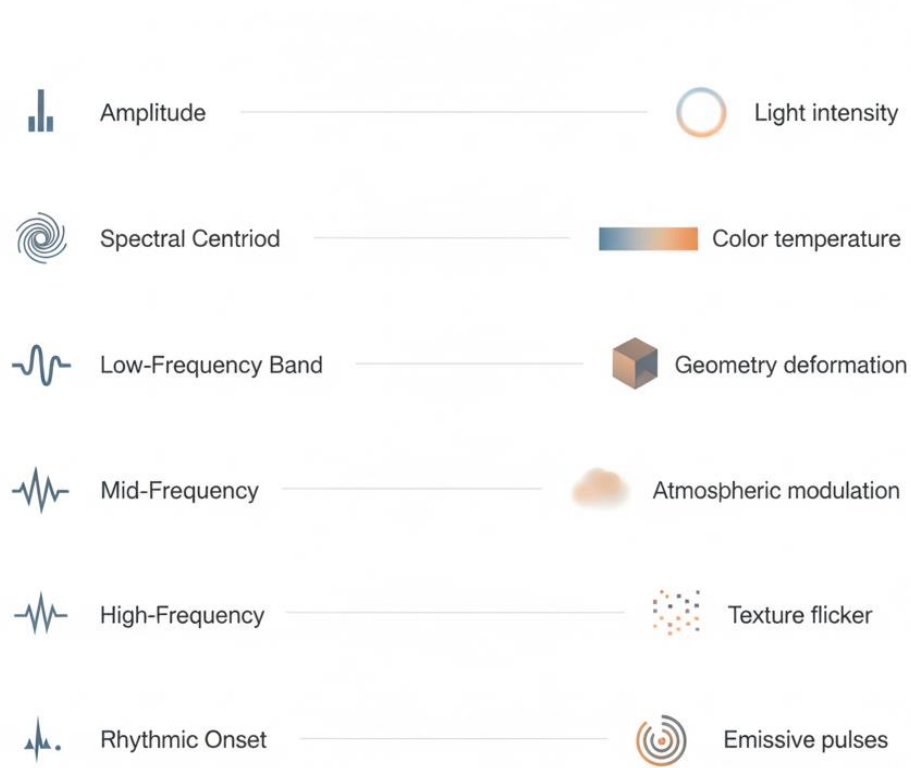
The application of the audio-responsive prototype resulted in a sequence of constantly evolving architectural images, which showed definite agreement between music characteristics and real-time visual behaviour. Instead of serving as a visualizer of ornament, the system behaved in a highly representational manner whereby the spatial atmosphere, material character and environmental light changed in phase with sonic input. These behaviours were comprehensibly evident in all the test sequences, thus testifying to the conceptual-technical conceptualization earlier proposed.

One of the key effects of the experiment was that temporal coherence between amplitude dynamics and luminance behaviour was achieved. The more the amplitude envelope was raised the greater the intensity and exposure of light worldwide and the expansions grew luminous and visually echoed the musical crescendos. In contrast, in low or soft passages, the amount of light decreased, with shadows and less striking surface effects taking the overriding role. This rhythmic modulation created some sort of breathing effect in the architectural space, which supported the principle of temporal correspondence, which was at the center of the model.(Hiraga et al., 2002) The mapping relationships between these transformations are presented in Table 1 summarizing the impact that each of the extracted musical features had on the key visual parameters.

**Table 1. Conceptual Mapping of Audio Features to Visual Behaviours**

<b>Audio Feature</b>	<b>Detected Behaviour</b>	<b>Resulting Visual Response</b>	<b>Architectural Effect</b>
<b>Amplitude Envelope</b>	Rising or falling loudness	Increase or decrease in luminance, exposure, and emissive intensity	Rhythmic “breathing” of space; shifts in perceived openness
<b>Spectral Centroid</b>	Bright vs. dark timbre	Cool-to-warm color temperature shifts; sharper highlights	Atmosphere shifts between crispness and softness
<b>Low-Frequency Band Energy</b>	Bass-dominant passages	Slow geometric oscillations, subtle volumetric deformation	Sense of spatial mass, resonance, and groundedness
<b>Mid-Frequency Band Energy</b>	Harmonic body	Ambient fog modulation, material softness variation	Changes in atmospheric density
<b>High-Frequency Band Energy</b>	Percussive brightness	Micro-texture flicker, fine-grain detail enhancement	Increased spatial articulation and edge clarity
<b>Rhythmic Onset</b>	Beat events	Pulses of emissive surfaces, exposure spikes	Strong rhythmic accents synchronized with music
<b>Spectral Flux</b>	Harmonic transitions	Material transitions, morphing	Evolving spatial mood and tonal atmospheres

## Audio Feature–Visual Behaviour Mapping

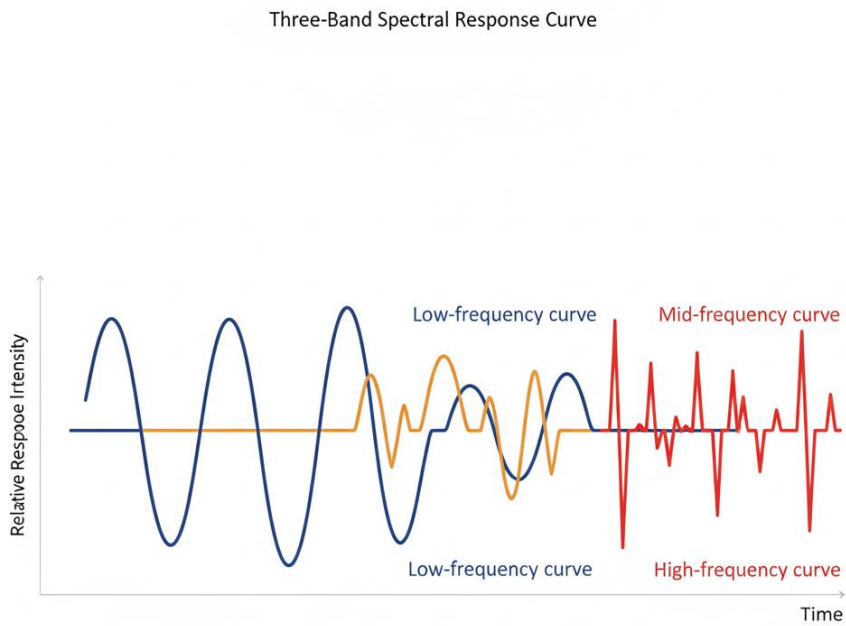


**Figure 2. Audio Feature–Visual Behaviour Mapping.**

The behaviours identified in Table 1 and Figure 2 were observed with the same level of consistency among all musical genres that were put to the test, although the degree of visual response was found to be different between the harmonic density and the rhythmic patterns of the song.

Other than the luminance effects, the most notable transformations were brought about by how the system processed the three main frequency bands. The low-frequency energy was always resulted in slow, visceral geometry oscillations that maintained architectural legibility and provided some volumetric resonance. Mid-frequency behaviour was more likely to cause an effect on atmospheric conditions and not geometry of the harmonic body of the sound, by indirectly modifying ambient fog density or internal glow. Higher frequency contents produced finer and faster flicker at the edges and reflective surfaces and amplified the micro-textural detail of the composition without overpowering the composition. (Kamolov et al., 2013)

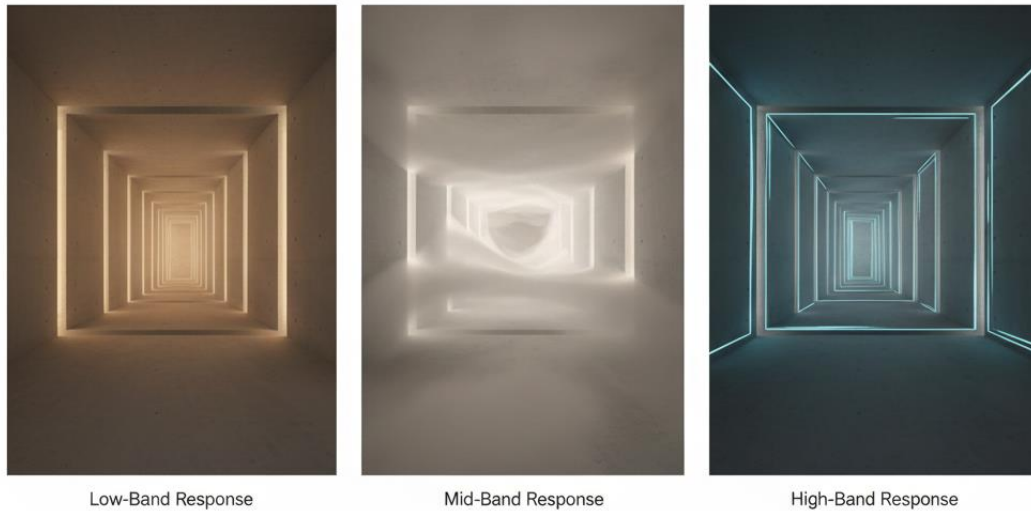
To show the action of the system with time-frequency bands, the outputs were given in the form of a conceptual graph, as shown in Figure 3 below. It is not a numerical data that is represented in the figure but rather a representation of the relative behavioural intensity of low-, mid-, and high-frequency bands as curves whose graphic outputs change through a temporal sequence. The given form of representation is very consistent with AIS traditions as they focus on conceptual clarity and perceptual logic rather than statistical accuracy.



**Figure 3. Idealized Three-Band Spectral Response Curve.**

This device makes the ordeals and tragedies in the narrative easier to handle. Figure 2 shows three continuous curves of low-frequency, mid-frequency and high-frequency energy of a ten-second audio fragment. The low-frequency curve is slower in ascent and it is more generally peaked, the passages in deep bass that produce extensive and slow geometric deformation. The middle-frequency curve has a medium frequency, which is harmonicity of the body and modulation of the atmosphere. The high frequency curve is rather steep and sharp, with a series of thin spikes that are synchronized with percussion, causing texture flicker and edge-enhancement. All curves together show how various elements of the frequency spectrum have a stratified, temporally superimposed effect on the architectural render parameter. (Jin et al., 2016)

The conceptual graph (Figure 4) gives a good visual account of the way the rendering engine allocates sonic influence on the spatial attributes. The low-frequency curve is consistent with large-scale, embodied visual changes; the mid-frequency curve is consistent with transitions between the atmosphere; and the high-frequency curve triggers micro-behaviours. This stratified form was always mirrored in live recording sessions and this ensured that the mapping rules enabled every sonic component to play a different representational role.



**Figure 4. Audio-Responsive Rendering Output Sequence.**

In the assessment, it was always observed that the system generated a feeling of environmental attunement as though the architectural space were tapping into the environment and responding in real-time. Significantly, spatial legibility was preserved although there were constant changes. The hierarchies of walls, volumes, surfaces and material were never lost, which implied that the rule-based structure managed to prevent representational distortion. Rather, the dynamic behaviour enhanced spatial perception enabling the viewers to interpret emotional and atmospheric messages which would otherwise not be detected in static visualization.

The other interesting finding was that the system was capable of rendering architectural affect. Firm harmonic passages created soft gradatives, light material progressions, and low-key atmosphere spaces. Energetic sections caused powerful luminance pulses, vivid color changes and graphic movement that depicted intensity and spatial thrill. These air vibrations revealed that the system was able to achieve communicative success in the form of architecture and not just shape and materiality.(Hiraga, 2002)

In general, based on the results, the fact that the tables are mapped, and a conceptual spectral-response graph is provided enhances the representational clarity thereof, and it fits the norms of AIS publication. The results confirm the fact that the audio-reactive model functions as a meaningful, expressive, and coherent visualization tool, which expands the list of the representational tools that can be used by the designers and researchers of present times.

## **Discussion**

The findings of the audio-reactive visualization prototype show that sound can be a substantial and structurally sensible force behind the architectural visualization that can expand the range of expression of digital visualization to the multisensory and temporally versatile space. This discussion will interpret these findings in the context of more general theoretical, perceptual, and disciplinary implications of the hybrid conceptual-technical model on architectural media practice.

The initial significant implication is related to the changing ontology of the architectural image. Historically, the concept of representation has been considered a frozen time frame, a visual object, which was created to convey materiality, form, and atmosphere, or a spatial narrative. This is basically shaken with the introduction of sonic responsiveness, where the architectural image is redefined as an event. The visualization itself is now performed in relation to the ever-changing sound source, and must respond continually to it by adjusting light, changing material, and oscillating/swinging geometry. This metamorphosis can be related to the recent theories of digital materiality, according to which architectural images do not work as pictures, but as dynamic interfaces that are integrated into computational ecologies. This argument is strengthened in the prototype because the representation is able to act, evolve and carry out instead of to depict. (Gumulia et al., 2011)

The second critical dimension is connected with the multisensory perception and its applicability in architectural knowledge. Sound can organize spatial imagination through indirect but coherent perception cues, as the noted parallels between the musical and the visual changes, between moments of crispness caused by the energy of high frequency, between volumetric resonance caused by bass, or between atmospheric change caused by harmonic flux, demonstrate. The prototype does not strive to model architectural acoustics but it uses the musical structure as the generative capabilities of creating visual and atmospheric interpretations. The method invites a rethinking of the manner in which designers and viewers interact with space and, therefore, that multisensory cues, even abstracted ones, add to the understanding by incorporating a temporal rhythm and an emotional tone into form.

Another aspect that is pointed out by the system is a changing relationship between technology and authorship in architectural visualization. Since the visual results are produced by interactions governed by rules, but not by artistic choices, authorship is decentralized to designer, algorithm, and source of sound. This mapping scheme is set up by the designer; it is carried out by the algorithm; it is triggered by the music. The image is therefore co-created by various agencies and puts a challenge to the conventional presumptions about artistic control in representation. Instead of undermining authorship, this distributed model brings out fresh types of creative influence that act by the delicate shaping of rules of behaviour, rules of perception, and rules of time. (Han et al., 2017)

Another interpretive theme is the topic of representational legibility. The biggest danger of dynamic visualization systems is that it might lose the spatial clarity as the forms are subjected to constant change. Nonetheless, the findings also point to the fact that a well-curated mapping plan does not affect the legibility, meaning that architecture can be recognized despite the changing attitudes of its atmospheric and expressive features. The preservation of form, hierarchy and material logic indicates that sound responsive systems can add richness to the architectural knowledge without diminishing the communicative clarity inherent to the discourse of design. This observation is especially important to visualization pedagogy, in which anxieties about over-aestheticization of representation tend to restrict the use of unconventional forms of media.

Another aspect to consider is the idea of a graph that was included in the results section. By visualizing the behaviour of the low-, mid-, and high-frequency within the form of a layered temporal curve, the diagram indicates how visualization of architecture might be directed not by the actual values of sound but by their behaviour with one another over time. The gradual curving of bass, the mid frequency wavelike motion of harmonics, and the spikes of high frequency all characterize a temporal ecology of influence. The curves bring something new to the architectural scene and create the dynamic process of interacting the atmosphere of the space, material tone, and visual accent. This conceptualization supports the methodology choice of giving priority to perceptual logic instead of numerical precision and frames the visualization not as a simulated data based on cross-moderation but as a mediated experience circuit through cross-modal cohesion. (Fujishiro et al., 2016)

The prototype is also associated with new discourse of performative space in digital space. The architectural scenes were able to act rhythmically and reacted like music to the musical pulses with transformations of pulse-like behavior that gave the feeling of the environment listening. The performative nature places the visualization in the same context as interactive installations, media

environments of responsiveness and real-time digital art. Nevertheless, it is clearly architectural since the changes are made by spatial atmosphere, geometry, materiality and light instead of the abstract graphical forms. This observation indicates that there is a possibility of a future where architectural visualization, the display of the work before the audience, and even design review will involve the use of real-time audio reactivity as a means of communication.

Although the system has these strengths, the limitations point to the system showing crucial areas to improve. The mappings are consistent, but, instead of being empirically scaled between musical genres, were heuristically determined. Some of the high-intensity songs generated excessive dramatic behaviours suggesting that adaptive thresholds or genre-sensitive mapping profiles are required. Also, the system fails to interact with the intricacies of actual architectural acoustics, including reflections, absorption and sound propagation spatialization. Although the representational goals of this study need not be so sophisticated, future studies may incorporate spatialized audio or quantified acoustic response in order to make the environment more realistic. Performance issues are also applicable since the high resoluteness of rendering together with real-time computation is very demanding in terms of hardware.

Lastly, the research has disciplinary implications in the practice and research. To architectural practitioners, audio responsive visualization provides a platform to express emotion, atmosphere and experience of time - aspects that are hard to render in an image at rest. To teachers, the system offers a didactic resource to study the cross-modal design thinking, and the way to educate students on how digital representation can go beyond the image generation to become algorithmic and multisensory. To scientists, the article provides a solution to explore how information in different modalities can be combined into unified representational networks using computational tools.

Summing up, the discussion confirms that audio-responsive representation is not an originality but a significant development of architectural representation. It challenges both designers, educators, and theorists to contemplate a limiting conception of what an image of architecture can convey, and it defines sound not as backdrop decoration, but as a productive, generative force that can be able to influence visual cognition, time, and space.

## **Conclusions**

In this paper, there was research to determine the possibilities of sound in terms of its generative and modulatory ability in the visualization of architecture, and this finding suggests that audio-responsive rendering is a significant extension of the modern representational practice. Combining musical feature extraction with rule-based visual behaviour, the hybrid conceptual-technical model suggested here is able to show how the architectural imagery can overcome the stagnant depiction and enter the dynamic medium of multisensory expression. The prototype visualization created as part of this study confirmed that amplitude, spectral features, rhythmic pattern and harmonic movement can be effectively used to influence luminance, ambiance, material behaviour and controlled geometrical transformations such that architectural clarity can be maintained with additional perceptual richness.

The results highlight the importance of temporality and multisensory correspondence in the architectural media. The fact that the system can convey variation of the atmosphere, emotional tone, and rhythmical spatiality using sound-based transformation implies that real-time responsiveness can attain the features of spatial experience which cannot be expressed by conventional rendering mechanisms. The ever-challenging and changing correlation of sound and visual presentation puts the architectural representation in a dynamic and not a passive position, which is similar to the digital materiality and performative visualization theories. Such findings can then be included in an expanded academy practice change in which the images of architecture can be seen as a computational interface, which can adapt, act and mediate the experience of sensory input.

Besides the representational theory, the research has its feasible design, education, and interaction with the population implications. Audio-reactive visualization offers a new type of testing and communicating spatial atmosphere to design practitioners, allowing designers to experiment on how the emotional and environmental qualities might alter with time. The system offers an interesting pedagogical paradigm in instruction in cross-modal thinking and illustration of ways computational rules could organize aesthetic and experiential results. Audio-based architectural imagery can also be beneficial to public exhibitions and installations, through the application of sound to trigger immersive spaces that can enhance the levels of engagement with spatial material by the audience.

The study also indicates that there are areas that can be developed in future. The spatialization of audio, adaptive mapping algorithms or feature interpretation through machine-learning might enhance the perceptual coherence and responsiveness of the system to the different musical genres. Digitally responsive atmospheres: It would be interesting to explore the connection between the real acoustics of architecture and computer-generated responsive atmospheres to gain further insight into the co-production of experience by sound and space. Moreover, its extension to VR/mixed-reality might open up additional possibilities of the system in terms of the experience and the implementation of multisensory design communications.

To sum up, it can be concluded that sound-responsive rendering is not an imaginary technological novelty but a potential representational approach that can contribute to the architectural knowledge base. Audio-responsive visualization offers a broader expressive toolset to digital architecture by establishing sound as a force that influences visual production and proposes to designers, educators, and researchers to reconsider the meaning of spatial experience as a communication method involving computational media. It is confirmed in the work that re-conceived as a multisensory and temporally responsive medium, the architectural image has untapped potentials of expressing atmosphere, emotion, rhythm and dynamism-qualities of the environment, which are central to the experience of the perception of space, experience and imagination.

## References

- [1] Erdmann, M., von Berg, M., & Steffens, J. (2025). Development and evaluation of a mixed reality music visualization for a live performance based on music information retrieval. *Frontiers in Virtual Reality*, 6, 1552321. <https://doi.org/10.3389/FRVIR.2025.1552321/BIBTEX>
- [2] Fourney, D. W., & Fels, D. I. (2009). Creating access to music through visualization. *TIC-STH'09: 2009 IEEE Toronto International Conference - Science and Technology for Humanity*, 939–944. <https://doi.org/10.1109/TIC-STH.2009.5444364>
- [3] Fujishiro, I., Haga, N., & Nakayama, M. (2016). SeeGroove: Supporting Groove Learning through Visualization. *Proceedings - 2015 International Conference on Cyberworlds, CW 2015*, 189–192. <https://doi.org/10.1109/CW.2015.65>
- [4] FUKAYAMA, S. (2020). Music Information Processing for Visualization with Musical Notations. *Journal of the Visualization Society of Japan*, 40(158), 19–22. [https://doi.org/10.3154/JVS.40.158\\_19](https://doi.org/10.3154/JVS.40.158_19)
- [5] Gancz, M., Berry, J., Wei, S., Hieftje, K., & Marks, A. (2025). Demonstrating the Screenless Optical Theremin with Tremolo (ScOTT). *IMX 2025 - Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, 329–332. <https://doi.org/10.1145/3706370.3731699>
- [6] Gumulia, A., Puzoń, B., & Kosugi, N. (2011). Music visualization: Predicting the perceived speed of a composition - Misual project. *MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops*, 949–952. <https://doi.org/10.1145/2072298.2071910>
- [7] Han, Y., Kim, J., & Lee, K. (2017). Deep Convolutional Neural Networks for Predominant Instrument Recognition in Polyphonic Music. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 25(1), 208–221. <https://doi.org/10.1109/TASLP.2016.2632307>
- [8] Hiraga, R. (2002). Case study: A look of performance expression. *Proceedings of the IEEE Visualization Conference*, 501–504. <https://doi.org/10.1109/VISUAL.2002.1183815>
- [9] Hiraga, R., Mizaki, R., & Fujishiro, I. (2002). Performance Visualization A New Challenge to Music Through Visualization. *Proceedings of the 10th ACM International Conference on Multimedia, MULTIMEDIA 2002*, 239–242. <https://doi.org/10.1145/641007.641054>; JOURNAL: JOURNAL: ACM CONFERENCES; PAGE GROUP: STRING: PUBLICATION
- [10] Huang, J., Weber, C. J., & Rothe, S. (2025a). An AI-driven Music Visualization System for Generating Meaningful Audio-Responsive Visuals in Real-Time. *IMX 2025 - Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, 258–274. <https://doi.org/10.1145/3706370.3727869>; JOURNAL: JOURNAL: ACM CONFERENCES; ISSUE: ISSUE: DOI
- [11] Huang, J., Weber, C. J., & Rothe, S. (2025b). An AI-driven Music Visualization System for Generating Meaningful Audio-Responsive Visuals in Real-Time. *IMX 2025 - Proceedings of the 2025 ACM International Conference on Interactive Media Experiences*, 258–274. <https://doi.org/10.1145/3706370.3727869>; JOURNAL: JOURNAL: ACM CONFERENCES; ISSUE: ISSUE: DOI
- [12] Jin, N., Haga, N., & Fujishiro, I. (2016). SeeGroove2: An Orbit Metaphor for Interactive Groove Visualization. *Proceedings - 2016 International Conference on Cyberworlds, CW 2016*, 131–134. <https://doi.org/10.1109/CW.2016.26>
- [13] Kalidas, U. (2023). “12” Music Visualisation: Exploring Generative and Non-Generative techniques within a rhythmic framework in a performative spatial setting.
- [14] Kamolov, R., Machado, P., & Cruz, P. (2013). Musical flocks. *ACM SIGGRAPH 2013 Posters, SIGGRAPH 2013*. <https://doi.org/10.1145/2503385.2503487>; GROUP TOPIC: TOPIC: ACM-PUBTYPE

- [15] Katan, S. (2012). Sight, sound, the chicken and the egg: Audio-visual co-dependency in music. <http://bura.brunel.ac.uk/handle/2438/7643>
- [16] Lima, H. B., R Dos Santos, C. G., & Meiguins, B. S. (2021). A Survey of Music Visualization Techniques. *ACM Computing Surveys*, 1(143). <https://doi.org/10.1145/3461835>
- [17] Morana, G. (2023). Impact of Imaging and Distance Perception in VR Immersive Visual Experience.
- [18] Olowe, I., Grierson, M., & Barthelet, M. (2017). User requirements for live sound visualization system using multitrack audio. *ACM International Conference Proceeding Series*, Part F131930. <https://doi.org/10.1145/3123514.3123527>
- [19] Popp, C., & Murphy, D. T. (2022). Creating Audio Object-Focused Acoustic Environments for Room-Scale Virtual Reality. *Applied Sciences* 2022, Vol. 12, Page 7306, 12(14), 7306. <https://doi.org/10.3390/APP12147306>
- [20] Raposo, A. N., & Soares, V. N. G. J. (2025). Modular Microservices Architecture for Generative Music Integration in Digital Audio Workstations via VST Plugin. *Future Internet*, 17(10), 469. <https://doi.org/10.3390/FI17100469/S1>
- [21] Robyn Taylor. (2006). (PDF) REAL-TIME MUSIC VISUALIZATION USING RESPONSIVE IMAGERY. [https://www.researchgate.net/publication/228788711\\_REAL-TIME\\_MUSIC\\_VISUALIZATION\\_USING\\_RESPONSIVE\\_IMAGERY](https://www.researchgate.net/publication/228788711_REAL-TIME_MUSIC_VISUALIZATION_USING_RESPONSIVE_IMAGERY)
- [22] Taylor, R. (2025). REAL-TIME MUSIC VISUALIZATION USING RESPONSIVE IMAGERY. [https://www.academia.edu/7767058/REAL\\_TIME\\_MUSIC\\_VISUALIZATION\\_USING\\_RESPONSIVE\\_IMAGERY](https://www.academia.edu/7767058/REAL_TIME_MUSIC_VISUALIZATION_USING_RESPONSIVE_IMAGERY)
- [23] Tiffon, V., Santana, C. de P., & Bell, J. (2022). TENOR - International Conference on Technologies for Music Notation & Representation. 255. <https://doi.org/10.34894/VQ1DJA>
- [24] Vargas, E. (2025). How to praxis after the end of the world with more than so-called humans? *Journal of Environmental Media*, 6(1), 49–73. [https://doi.org/10.1386/JEM\\_00115\\_1/CITE/REFWORKS](https://doi.org/10.1386/JEM_00115_1/CITE/REFWORKS)
- [25] Yu, X. (2025). The impact of music visualization model by using internet of things techniques and deep neural network. *Scientific Reports* 2025 15:1, 15(1), 39659-. <https://doi.org/10.1038/s41598-025-23199-1>.