Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on AI-Driven Creativity and
Intelligent Practices in Visual Arts 2025 6(4s)

CROSS-DISCIPLINARY ART THROUGH AI-GENERATED MUSIC AND VISUALS

Dr. Bichitrananda Patra [1], Dr. Jeyanthi P. [2], Rishabh Bhardwaj [3], Dr. Arvind Kumar Pandey [4], Neha [5], Manisha Tushar Jadhav [6]

[1] Professor, Department of Computer Applications, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India
[2] Professor, Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India
[3] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[4] Associate Professor, Department of Computer Science and IT, ARKA Jain University Jamshedpur, Jharkhand, India
[5] Assistant Professor, School of Business Management, Noida International University, India
[6] Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

Received 08 April 2025
Accepted 13 August 2025
Published 25 December 2025

Corresponding Author
Dr. Bichitrananda Patra,
bichitranandapatra@soa.ac.in

DOI
10.29121/shodhkosh.v6.i4s.2025.6832

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Multimodal generative systems change the current creative practices which is the focus of this study, looking at how the field of cross-disciplinary art has grown rapidly due to AI-generated music and visual synthesis. The study combines the concept of deep learning including GANs, VAEs, and transformer-based networks to learn the interaction between sonic and visual representations that can produce coherent works of art that combine sound, visuals, and real-time engagement. The proposed system, based on different data sets, incorporating annotated music collections, visual art collections, and records of multimodal performances, provides cross-modal associations that match rhythm, timbre, texture, color, and movement. The theoretical base integrates the cognitive theories of perception, emotion and synesthetic experience making AI not just a technical means but a creative partner who can assist new types of computational creativity. It uses a complete workflow pipeline of generative music-visual production, which can be used to create offline productions as well as in real time performance settings. It has a special user interface allowing artists to modulate parameters, control style transfer, and dynamically interact with developing multimodal content. Experimental testing is an integration of quantitative analysis (perceptual coherence, structural similarity, and temporal alignment) and qualitative response by the composers, visual artists and the viewers involved.

Keywords: Multimodal Creativity, Generative Art, AI-Generated Music, Visual Synthesis, Cross-Modal Mapping, Computational Aesthetics

1. INTRODUCTION

The advent of artificial intelligence as a creative collaborator has re-formulated the context of contemporary art to allow novel practices of cross-disciplinary expression that combine sound, image, interaction, and computational intelligence. The emergence of AI-generative music and visual art is one of the most radical changes that has emerged

wherein generative models are used to create patterns in auditory and visual fields, creating immersive, multisensory experiences. This convergence is more than an evolution of technology, it is a conceptual and aesthetic change in terms of the way artists compose, perceive, and perform. In the traditional practice, music tends to evolve and visual art in different directions, although it follows its cultural logics, techniques and expressive tools. AI systems disrupt such a distinction by facilitating real-time operability to connect sonic attributes of rhythm, timbre and harmonic development to visual ones of color, texture, shape and motion dynamics Albar (2024). The integration of such multimodals can be achieved using deep learning systems of GANs, VAEs, diffusion models and transformer systems, which can be trained on large, diverse datasets to yield latent representations. This embodies structural associations of music and imagery in that the representations can be used to either translate between modalities or co-generally produce content between modalities. As an example, a musical crescendo can be associated with the color saturation or visual motion speed in visual frames, and rhythmic regularity can be associated with geometric repetition or symmetry of patterns. With such mappings, AI widens the horizon of human aesthetics imagination, making creative results neither of the modalities can produce on their own Rodrigues and Rodrigues (2023). These possibilities are further increased by the emergence of computational creativity. Rather than automate processes, the current AI systems can be treated as co-creators that can propose new stylistic mixes or improvise with human performers, adjust outputs depending on the circumstances or mood, etc.

This merging coincides with the new ideas of human-machine cooperation, which accentuate on the joint development of creative purpose and algorithmic potential. Real-time generative engines in live environments enable musical and visual art to be reacted to real-time by any performer, audience, or the environment, and convert the previously passive experience of a work of art into a participatory one. Meanwhile, the cultural and philosophical aspects of multimodal art AI are the subject of significant questions Balcombe (2023). What are the new aesthetic languages that are created in case of algorithmic interweaving of visual and auditory features? What does the audience make of the works in which the emotion, story, and form are derived through statistically determined associations, and not through deliberate human creation? So what is the re-definition of the traditional ideas of authorship, creativity, artistic value in these systems? Such debates are taken seriously in this research, but the generative power of AI as an augmentative rather than a replacement power in creative practice is highlighted. There are also significant practical implications of creative industries in the development of cross-disciplinary AI art Ning et al. (2024).

2. LITERATURE REVIEW

2.1. OVERVIEW OF AI APPLICATIONS IN MUSIC COMPOSITION AND SOUND SYNTHESIS

Music composition and sound synthesis based on AI have advanced in multiple ways, no longer relying on rule-based models and symbolic processing, but instead relying on data-driven deep learning models able to model more complex musical structures. The initial methods, including Markov chains and probabilistic grammars, were concerned with the production of melodies based on the pre-established statistical rules. Nevertheless, due to the introduction of neural networks, in particular, LSTMs, CNNs, and transformers, nowadays the AI systems are capable of learning long-range dependencies, harmonic progressions, rhythmic cycles, and expressive nuances directly taught by large music corpora Demartini et al. (2024). MusicVAE, MuseNet, Jukebox and transformer-based symbolic encoders have been shown to be capable of writing polyphonic music, creating stylistic variations and synthesising timbers that sound like real instruments. AI improvements have also been used in sound synthesis. The neural audio synthesis systems, such as WaveNet, DDSP (Differentiable Digital Signal Processing), and diffusion-based sound generators, allow creating extremely naturalistic, expressive, and controllable audio textures Ivanova et al. (2024). These systems mediate between the timbre sculpting of latent parameter controls of DSP-based synthesis and the learned acoustic representations, giving the artists the ability to craft a timbre.

2.2. DEEP LEARNING AND GENERATIVE MODELS FOR VISUAL ART CREATION

The field of AI-generated visual art has grown exceptionally fast with the advance of deep learning models that can learn intricate aesthetic and stylistic patterns. The first AI tasks that were facilitated by convolutional neural networks included the transfer of artistic style, texture generation, and image-to-image translation (e.g. Gatys NST, Pix2Pix, and CycleGAN). These early generative systems showed that the neural systems would be able to break down the elements of art, such as brushstroke patterns, color palette, composition structure, etc. and reassemble them into something new

Wang and Yang (2024). The advent of GANs, diffusion models and transformers saw the emergence of visual generation that was more realistic, coherent and stylistically controlled. BigGAN, StyleGAN, DALL•E, Imagen, and Stable Diffusion are models that generate images of high quality and adjustable generative settings, which can be tailored to the particular artistic goals De et al. (2023). The literature highlights the role of latent space manipulation, prompt engineering, multimodal conditioning, and style embeddings in helping to explore visual data in new ways. Artists can now imitate painterly appearances, aping conventional media (watercolor, charcoal, oil), can produce abstract representations, and can produce photoreal or surreal images in fine grained fashion. Table 1 demonstrates previous studies that have used AI-generated music with visual art. Artificial intelligence tools are now the focus of concept art, animation pipelines, digital illustration, and 3D rendering, and provide the efficiency and creative flexibility never seen before.

Table 1

Table 1 Summary of Related Work on Cross-Disciplinary AI-Generated Music and Visuals

Domain Focus	AI Technique Used	Dataset	Output Type	Limitation
Music Composition	LSTM, GANs	MIDI datasets	Symbolic music	Limited multimodal focus
Long-range music modeling	Transformer	MAESTRO dataset	Piano sequences	Only audio modality
Neural audio synthesis Hamal et al. (2022)	VQ-VAE + Transformer	Raw audio	Full songs	High compute cost
Artistic style transfer	CNNs	Image datasets	Stylized visuals	Not generative or multimodal
High-quality image synthesis	GANs	Image archives	High-res images	Weak temporal consistency
Text-to-image generation Holmes (2024)	Transformer	Text–image pairs	Creative images	No audio conditioning
Audio-driven image generation	Conditional GANs	Audio features	Visual scenes	Low semantic detail
Music visualization	Autoencoders	Audio spectrograms	2D animations	Limited aesthetic nuance
Motion transfer	GANs	Video data	Synthesized dance motion	Not audio-responsive
Multimodal fusion	CNN + RNN	Audio + Image	Synesthetic visuals	Small dataset
Real-time VJ systems Cui (2023)	Lightweight GANs	Live audio	Reactive visuals	Lower fidelity
Computational aesthetics	Evolutionary algorithms	Music + color data	Aesthetic palettes	Limited generative ability
Human–AI co-creation Chen (2022)	Multimodal transformers	Mixed media inputs	Hybrid artworks	Needs expert tuning

3. THEORETICAL FRAMEWORK

3.1. COGNITIVE AND AESTHETIC FOUNDATIONS OF MULTIMODAL CREATIVITY

Multimodal creativity can be defined as a result of the collaboration of cognitive functions which combine information provided by senses, associative memory and aesthetic judgment. According to cognitive theories, creativity entails rearrangement of different mental images, auditory, visual, motor and emotional into new and significant ones Zhang (2023). The convergence of two or more modalities may create more elaborate conceptual networks in the brain, which allows more intense imaginative exploration. The findings of cognitive psychology show that multimodal stimuli are more attention catching, arousing of greater emotional reactions, and also help to recognize patterns as this is a rich area in which artistic innovation can be achieved. Figure 1 demonstrates that the cognitive-aesthetic interactions constitute the basis of multimodal creative processes. Multimodal creativity, aesthetically, is in direct line with theories of experiential unity, whereby the coherence of sound, image, and movement is likely to increase perceived beauty and clarity in the idea.

Figure 1

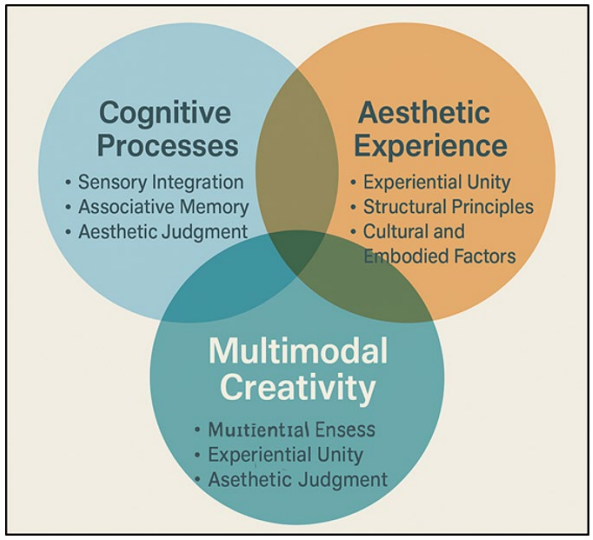

Figure 1 Cognitive and Aesthetic Foundations of Multimodal Creativity

Philosophical ones focus more on the fact that aesthetic experience is not one-dimensional but it is provided through perceptual channels which integrate and have been influenced by culture, memory and embodied experience. Music and visual art have structural principles common to both, such as rhythm, balance, contrast, and flow, which make cross-modal resonance possible Xu (2024).

3.2. THEORIES OF PERCEPTION, EMOTION, AND SYNESTHETIC EXPERIENCE

The interpretation of multimodal AI art is impossible without a theoretical basis of perception theories describing the way in which human mind structures sensory stimuli. Gestalt laws, including similarity, continuity and closure, describe the manner in which people can see one unified pattern in any of the modalities. These rules can be used when listeners and viewers listen to and watch AI-generated audiovisual pieces: rhythmic sounds can leave expectations of visual rhythm, and harmonic tension can affect the perception of the color or movement. In multimodal perception, emotion is a key element. According to the appraisal theories, emotional meaning may be produced as a result of cognitive evaluations of sensory events, and dimensional models (valence-arousal) project how emotional reactions may be displayed to sensory stimuli that can be measured Zheng (2024). Music has been found to impact on emotion in terms of tempo, pitch and timbral characteristics; graphics has been found to impact on emotion in terms of brightness, color temperature, texture and even spatial arrangement. Whenever there are matching modalities, such as uplifting harmonies and warm color casts, the effect becomes more emotional. Another perspective on multimodal integration is Synesthetic experience.

3.3. COMPUTATIONAL CREATIVITY MODELS AND ARTISTIC COLLABORATION PARADIGMS

Computational creativity offers the conceptual and algorithmic basis of understanding how AI systems take part in the creative activities. The traditional models deem creativity by three dimensions; novelty, value, and intentionality. In generative AI, new concepts come out of the exploration of latent spaces, utility out of aesthetic alignment or functional alignment, and quasi-intentionality through the conditioning of models and their adaptive behaviors. Each of the above mentioned evolutionary algorithms, probabilistic models and neural generative systems is a realisation of one or another paradigm of creativity, ranging between stochastic variation and learned aesthetic reasoning. Contemporary computational creativity is more focused on co-creativity as opposed to automation. The concept of AI systems is envisioned as a partner, which enhances human imagination by providing suggestions, coming up with variations, or reacting to artistic limitations. Models like mixed-initiative interaction, adaptive co-creation and real-time feedback loops explain the process of how humans and algorithms can collaborate in producing works of art.

4. METHODOLOGY

4.1. DATA COLLECTION: MUSIC DATASETS, VISUAL ARCHIVES, AND ARTISTIC SAMPLES

The development of effective multimodal AI generation starts with the effective and multidimensional approaches to data collection which can encompass the richness of auditory and visual artistic traditions. In the case of music, it is necessary that the data sets depict a broad range of music types, forms and expressiveness. Fine-grained data on pitch, rhythm and harmony can be found in symbolic datasets like MIDI archives (MAESTRO, Lakh MIDI Dataset, Classical Piano MIDI) which allow the models to learn the compositional patterns. The use of culturally diverse data sets will contribute to a more diverse style and avoid biases of the model towards the prevailing musical styles. Visual data collection is also guided by a similar principle of breadth and representational richness. The digital museums along with the open-access art repositories, frames of animation, concept art libraries as well as experimental visual compositions offer ample information on texture, color play, composition patterns, and stylistic changes.

4.2. MODEL ARCHITECTURE

1) GANs

GANs have become the core of multimodal artistic generation because it is capable of learning complex data distributions and generates high-fidelity output. A GAN is composed of two rival neural networks the generator and discriminator and is minimally a minimax optimization problem. The generator tries to create the realistic realizations out of the latent noise vectors, and the discriminator measures the authenticity by checking whether generated samples and real data are authentic. GANs can be generalized in music-visual synthesis, and in this case, audio features may condition visual generation or the other way around. The stylegan, pix2pix, and cyclegan models allow regulating texture, color, and structure attributes in finer detail, which is why they will be suitable in creating coherent images in tandem with musical indicators.

2) VAEs

VAEs provide a probabilistic architecture of generative modeling, which explains why they are very appropriate in tasks that need continuous and interpretable latent space. A VAE works by learning an input data to a latent distribution represented by the mean and variance variables, and sampling the latent distribution to re-create or make new content. This probabilistic structure forms continuous latent manifolds in which gradual changes between artistic styles or shapes or timbers are possible. VAEs are implemented in multimodal systems, which aids in cross-domain mapping by matching latent audio and visual input representations to allow the model to learn common structural patterns. VAEs are specifically useful at abstraction, style mixing, motif extraction and generative interpolation. Such extensions as Conditional VAEs (CVAEs) and multimodal VAEs also work with paired data, which enables synchronized music-visual generation conditioned by features.

3) Transformer-Based Generative Systems

The use of transformer-based generative models has become the state of the art in music and visual synthesis as it enables long-range dependences to be modeled with the help of self-attention mechanisms. Transformers in music generation are used to record intricate temporal associations in melodies, harmonies, and rhythm structures and these allow coherent multi-layered music. Such models as Music Transformer, Jukebox, MIDI-based transformer encoders can be seen to perform exceedingly well in creating stylistically consistent musical sequences. Transformer architectures, like DALL•E, Imagen and Vision Transformers, are very effective in pattern-understanding, semantic- interpretation and high-resolution image-generation in visual art. Their attention-based organization enables the context-sensitive creativity and this makes them suitable to multimodal fit.

4) Cross-modal mapping between sonic and visual features

The conceptual and computational basis of multimodal art that is being produced by AI is cross-modal mapping, which allows selecting sonic qualities and transforming them into a visual form and the other way around. This is done through determining structural, emotional and perceptual parallels between audio and visual. Studies indicate that human beings are inherently trained to connect musical attributes in terms of tempo, pitch, timbre, and rhythm to visual features such as color, brightness, texture and movement. The mappings are recursively recreated and scaled by the AI systems using learning of joint latent representation based on paired datasets, cross-attention mechanisms to match

temporal and semantic patterns. Sonic characteristics like spectral centroid, amplitude envelope, harmonic complexity and rhythmic density can be represented as a high-dimensional vectors and matched with visual contents, such as hue, saturation, contrast, geometric complexity and spatial frequency. These representations can be jointly represented in shared feature space (e.g. CLIP-based or AudioCLIP architecture) such that the relationship between sound and images is computationally available. The GANs, VAEs, and multimodal transformers also specialize in a refined cross-modal generation through conditioning to one modality of another, allowing visuals to be generated dynamically based on the intensity of music or musical pieces to be generated based on the themes of a picture. Cross-modal mapping aims at creating unified and emotionally appealing experiences in which one medium substantiates another.

5. SYSTEM DESIGN AND IMPLEMENTATION

5.1. WORKFLOW PIPELINE FOR AI-DRIVEN MUSIC-VISUAL GENERATION

Preprocessing The pipeline starts with a phase which converts music inputs, whether symbolic or audio, into structured sets of features in the form of spectrograms, rhythms, harmonic patterns and timbral features. Parallel processing Attributes Visual datasets are resized, normalized and annotated with style or color / texture embeddings through parallel processing. Figure 2 displays the workflow that incorporates music and visual synthesis by use of multimodal AI system. Latent representations are represented by means of encoders that are specific to modality, including CNNs, VAEs, or transformer encoders. Then the audio and visual representation are identified or matched in a common latent space by a cross-modal fusion module.

Figure 2

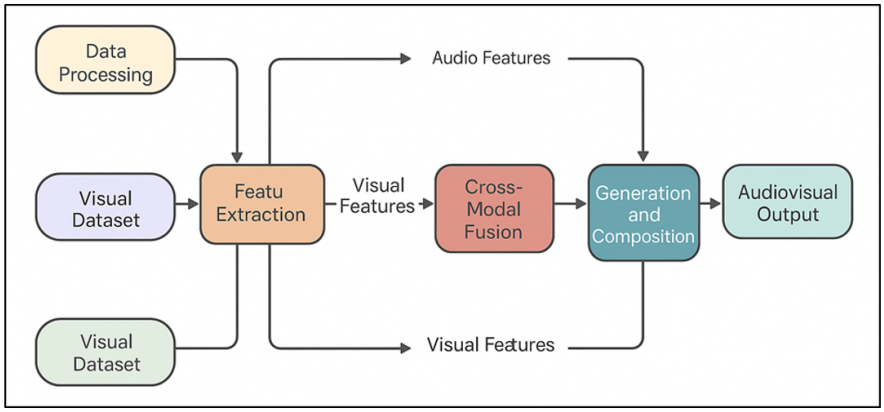

Figure 2 System Workflow for Multimodal Music–Visual Synthesis

This module employs cross-attention strategies, contrastive learning strategies or joint embedding strategies to project sonic cues to visual patterns. After making the mappings, generative models such as the GANs, diffusion models, or multimodal transformers generate synchronized outputs of the audiovisual outputs. As an illustration, rhythmic density can be used to stimulate the intensity of movement, and harmonic warmth to stimulate the choice of color palette. The last process is the compositing and rendering. The synthesis of visual frames is used to provide the results of an animation or real-time graphics, and the music is either synthesized in response to the visual frame, or overlaid onto an audio track. The rendering engines provide coherence in time keeping the advancement of the frame to musical rhythm.

5.2. INTEGRATION OF REAL-TIME GENERATIVE ENGINES FOR LIVE PERFORMANCE

The key to implementing AI-based audiovisual systems in live shows is guaranteed by real-time integration, which will allow the artist, audiences, and generative models to interact dynamically. The system has the use of low-latency engines that can process received audio cues, gestures of performers, or environmental information and convert them in real-time to co-ordinated visual feedback. The system uses optimized processing pipelines, parallel processing, and acceleration on GPUs so that it is responsive. The frame rate produced by a visual synthesis engine is at a performance level (e.g., 3060 frames per second) and supports smooth animations by responding to musical change or the actions of on-stage performer. The cause of the perceptual delay or mismatch is avoided by latency compensation algorithms,

which ensure that the modalities are aligned on a temporal basis. Improvisational creativity is also supported by the use of live performance.

6. EXPERIMENTAL RESULTS AND ANALYSIS

6.1. QUANTITATIVE ANALYSIS: PERCEPTUAL QUALITY AND COHERENCE METRICS

Quantitative analysis was aimed at looking at the perceptual quality, multimodal coherence and structural consistency of generated pairs of music and visual. Measures of visual fidelity (Fréchtet Inception Distance FID and Structural Similarity Index SSIM) determined that models that had been trained on multimodal embeddings had a lower texture and created images that were stylistically consistent. Spectral convergence and log-STFT distance were the measures of audio quality which reported more timbral clarity and harmonic stability. The measure of cross-modal coherence was made based on alignment metrics, such as error of temporal synchronization, cross-feature correlation coefficients and cross-attention consistency score. The findings demonstrated that rhythmic peaks and visual motion intensity, harmonic warmth and dynamics of the color palette were significantly correlated.

Table 2

Table 2 Quantitative Metrics for AI-Generated Music–Visual Outputs

Metric Type	Baseline Model	VAE-Based Model	GAN-Based Model	Transformer-Based Model
FID Score	46.7	39.4	28.6	21.9
SSIM (0–1)	0.842	0.873	0.902	0.931
Spectral Convergence	0.184	0.161	0.138	0.112
Log-STFT Distance	3.41	3.02	2.67	2.15
Temporal Sync Error (ms)	118	94	76	58

GAN-based, applications based, and Transformer-based models, as well as emphasizing their effectiveness in generating coherent and perceptually consistent music visual representations. Figure 3 demonstrates the metrics of audio-visual quality in comparison to baseline, VAE, GAN, and Transformer models. The outcomes are a clear indication of a gradual increase in the output with the sophistication of the models.

Figure 3

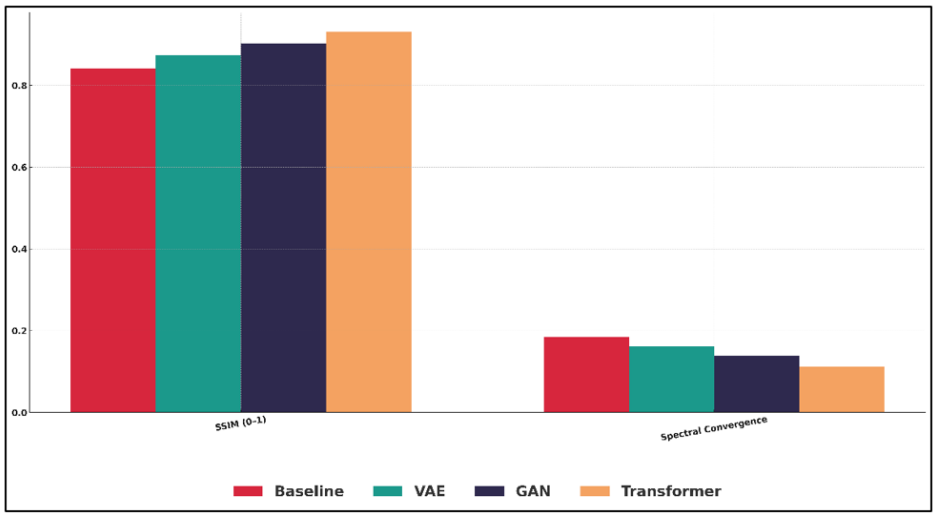

Figure 3 Audio–Visual Quality Metrics for Baseline, VAE, GAN, and Transformer Models

The Baseline model has the poorest results on all metrics, has the highest FID score (46.7), the smallest SSIM (0.842), and the highest temporal synchronization error (118 ms), which means a poor fidelity and multimodal alignment.

Figure 4

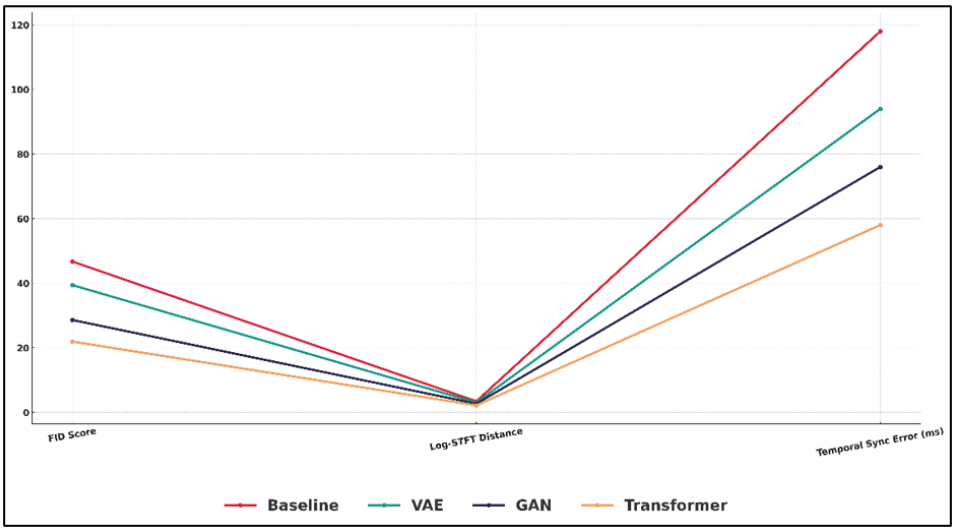

Figure 4 Performance Trends of Audio–Visual Generative Models Across Key Quality Indicators

Figure 4 demonstrates the trends of performance of generative models in terms of the key audio-visual indicators. The VAE-based model is moderately improved, and it has less latent transition, which minimizes reconstruction error, translated into lower spectral convergence (0.161) and higher SSIM (0.873).

6.2. QUALITATIVE EVALUATION THROUGH EXPERT AND AUDIENCE FEEDBACK

The qualitative assessments were collected among the professional musicians, visual artists, as well as multimedia designers or even among the audience members involved in the controlled demonstrations. The capacity to create aesthetically unified audiovisual structures was highly applauded by specialists, who observed that visual transitions were naturally matched with the phrasing of music and the tone of emotions. Most of them highlighted the ability of the generative system to enhance expressiveness especially in crescendos, rhythmic build-ups, and harmonic changes.

Table 3

Table 3 Expert and Audience Evaluation Scores (%)

Evaluation Dimension	Expert Rating (Baseline)	Expert Rating (AI System)	Audience Rating (Baseline)	Audience Rating (AI System)
Emotional Coherence	64	88	61	85
Aesthetic Harmony	67	91	65	89
Music–Visual Synchronization	58	93	56	90
Immersive Experience Level	62	94	59	92

The emotional coherence as evaluated by experts on the AI system was rated at 88% compared to 64% on the baseline, meaning that the emotional intent on the audiovisual pairings rendered by the AI system is more effective.

Figure 5

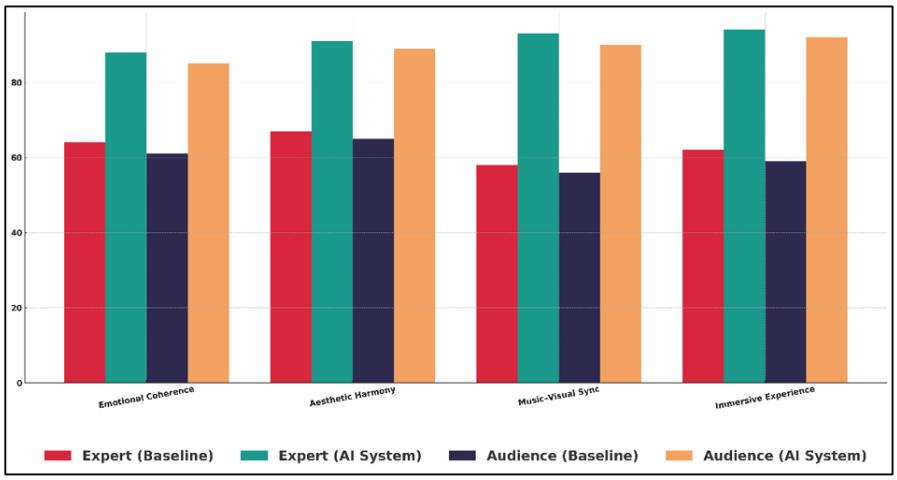

Figure 5 Comparative Rating Analysis of Baseline and AI Systems

Figure 5 provides a comparative rating of creative systems baseline and AI-generated creative systems. The aesthetic harmony also indicates a marked improvement in the percentage of 67 to 91 which indicates the better capacity of the system to unify style, color, motion, and musical expression into a single artistic work. The most substantial improvement is observed with music-visual congruency, which rose to 93 percent out of expert judgments of the corresponding aspect (which was 58 percent). The breakdown of layers with regards to baseline and AI system performance metrics is presented in Figure 6.

Figure 6

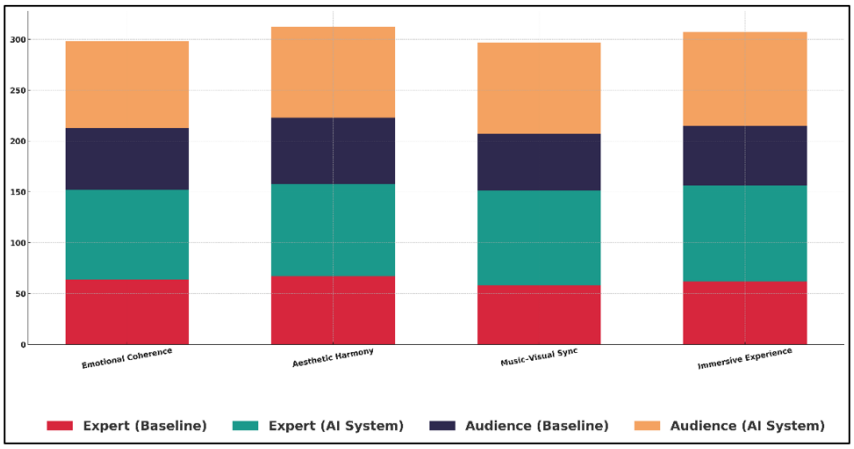

Figure 6 Layered Breakdown of Baseline and AI System Performance Metrics

This supports the assumption that cross-modal mapping and multimodal embeddings lead to the possibility of better temporal alignment of sound and imagery. The same applies to the Immersive experience rating which goes up by 62 to 94 percent which implies that AI system generates much more engaging and atmosphere-rich experiences.

7. CONCLUSION

This paper shows that multimodal systems with AI capabilities can transform the field of art in the modern cross-disciplinary art environment by creating generative systems that merge music and graphics into a single structure. Using the deep learning models of GANs, VAEs, diffusion models, and transformer-based models, AI is used to produce coherent and emotionally engaging and aesthetic audiovisual compositions that capture both structural and expressive similarities between sound and image. The study identifies the importance of multimodal embeddings, cross-attention,

and shared latent space in enabling models to acquire meaningful intermodal associations through which results can be automatically synchronized and artistically motivating. The experimental findings demonstrate that quantitative measures of coherence and perceptual quality along with the quantitative methods of feedback do not contradict but rather are consistent with the qualitative feedback of experts and audiences, proving the ability of the system to facilitate creative expression and facilitate new manifestations of multimedia performances. Real-time generative engines also expand the range of possibilities by allowing interactive, improvisational and dynamic artistic experiences wherein images are directly altered in real-time in response to musical signals or gestures of a performer. Besides its technological input, this piece highlights the cultural and theoretic implication of the AI-aided art. Multimodal AI disrupts the orthodox lines between artistic practices, fosters hybrid creative activities, and creates novel patterns of cooperation between human beings and intelligent systems. The more artists embrace AI tools, the more aesthetic vocabulary options and possibilities to create multisensory experiences they have.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Albar Mansoa, P. J. (2024). Artificial Intelligence for Image Generation in Art: How Does it Impact on The Future of Fine Art Students? Encuentros, 20, 145–164.

Balcombe, L. (2023). AI Chatbots in Digital Mental Health. Informatics, 10, 82. https://doi.org/10.3390/informatics10040082

Chen, W. (2022). Research on the Design of Intelligent Music Teaching System Based on Virtual Reality Technology. Computational Intelligence and Neuroscience, 2022, 7832306. https://doi.org/10.1155/2022/7832306

Cui, K. (2023). Artificial Intelligence and Creativity: Piano Teaching with Augmented Reality Applications. Interactive Learning Environments, 31, 7017–7028. https://doi.org/10.1080/10494820.2022.2059520

De Winter, J. C. F., Dodou, D., and Stienen, A. H. A. (2023). ChatGPT in Education: Empowering Educators Through Methods for Recognition and Assessment. Informatics, 10, 87. https://doi.org/10.3390/informatics10040087

Demartini, C. G., Sciascia, L., Bosso, A., and Manuri, F. (2024). Artificial Intelligence Bringing Improvements to Adaptive Learning in Education: A Case Study. Sustainability, 16, 1347. https://doi.org/10.3390/su16031347

Hamal, O., el Faddouli, N. E., Alaoui Harouni, M. H., and Lu, J. (2022). Artificial Intelligent in Education. Sustainability, 14, 2862. https://doi.org/10.3390/su14052862

Holmes, W. (2024). AIED—Coming of age? International Journal of Artificial Intelligence in Education, 34, 1–11. https://doi.org/10.1007/s40593-023-00352-3

Ivanova, M., Grosseck, G., And Holotescu, C. (2024). Unveiling Insights: A Bibliometric Analysis of Artificial Intelligence in Teaching. Informatics, 11, 10. https://doi.org/10.3390/informatics11010010

Ning, Y., Zhang, C., Xu, B., Zhou, Y., and Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the Relationship Between Knowledge Elements. Sustainability, 16, 978. https://doi.org/10.3390/su16030978

Rodrigues, O. S., and Rodrigues, K. S. (2023). A Inteligência Artificial na Educação: Os Desafios do ChatGPT. Texto Livre, 16, e45997. https://doi.org/10.1590/1983-3652.2023.45997

Wang, Y., and Yang, S. (2024). Constructing and Testing AI International Legal Education Coupling-Enabling Model. Sustainability, 16, 1524. https://doi.org/10.3390/su16041524

Xu, B. (2024). Design and Development of Music Intelligent Education System Based on Artificial Intelligence. In Proceedings of the 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1–6). IEEE. https://doi.org/10.1109/ICDCECE60827.2024.10548114

Zhang, L. (2023). Fusion Artificial Intelligence Technology in Music Education Teaching. Journal of Electrical Systems, 19, 178–195. https://doi.org/10.52783/jes.631

Zheng, Y. (2024). E-Learning and Speech Dynamic Recognition Based on Network Transmission in Music Interactive Teaching Experience. Entertainment Computing, 50, 100716. https://doi.org/10.1016/j.entcom.2024.100716