

End-to-End Music-Driven Stage Lighting: A Co-Creative Framework

Tobias Wursthorn¹[0000–0003–1471–5253]*, Michael Kohl¹[0009–0005–9351–5058],
Christof Weiß¹[0000–0003–2143–4679], Kai von Luck¹[0000–0002–5177–3025], Peer
Stelldinger¹[0000–0001–8079–2797], and Larissa Putzar¹[0009–0004–4902–5548]

Hamburg University of Applied Sciences, Hamburg, Germany
tobias.wursthorn@haw-hamburg.de

Abstract. Professional stage lighting operates in a high-dimensional control space, creating a substantial authoring burden that limits creative iteration. Fully scored shows synchronized to music remain a major production bottleneck, motivating assistive tools that augment rather than replace designer agency. This paper introduces an end-to-end music-to-lighting system built around a Dual Abstraction Layer that separates slow-varying aesthetic dynamics from fine-grained rhythmic mechanisms. A Diffusion model generates Global Intention trajectories capturing intensity, spatial contrast, and color across entire songs, while a Conformer predicts Segment-Based Oscillator patches that parameterize beat-aligned patterns within musically coherent sections. Trained on 1,400 minutes of professionally authored shows (280 songs), the system produces abstract proposals that an interpreter can retarget to arbitrary rigs and expose for co-creative editing. Mixed-method evaluation combines intention-based metrics with a formative expert study (N=2, each with more than 10 years of professional experience). The system achieves an Overall Quality Score of 76.4% relative to professional references, with experts highlighting strong rhythmic responsiveness alongside the critical need for ecosystem integration. To our knowledge, this is the first Automatic Stage Lighting Control system to combine rig-retargetability with semantic editing at the intention level.

Keywords: Automatic stage lighting control · Abstraction · Retargeting · Generative models · Co-creative systems

1 Introduction

Modern concert productions routinely employ dense arrays of intelligent luminaires controlled over DMX-512 [2] or related protocols. Large control systems expose tens to hundreds of thousands of controllable parameters per session [13]. This capacity enables intricate spatial, chromatic, and temporal compositions, but imposes significant authoring overhead. Designers must translate musical form and artistic intent into thousands of tightly synchronized cues (discrete

* Corresponding author

programming states triggered at specific moments) under time pressure, which often results in reduced structural complexity or reliance on busking (live, improvisational triggering of pre-built looks in real time) rather than fully scored, music-aware shows.

Automatic Stage Lighting Control (ASLC) aims to reduce this bottleneck, yet faces a representational gap between low-level control channels and high-level design concepts. Raw audio and DMX values lack explicit representation of high-level concepts, musical form (verse, chorus) or visual intent (warm wash, dramatic blackout). Existing ASLC approaches map handcrafted audio features directly to fixture parameters [14, 9] or learn end-to-end mappings to raw control channels [5, 20]. These strategies remain tied to specific rigs, yield outputs that resist editing, and obscure the relationship between musical intent and control data, limiting their value as assistive tools for professional workflows.

This work demonstrates that an explicit abstraction layer between raw control data and trained generative models is a practical basis for retargetable, editable generation. The system produces abstract proposals, semantically meaningful mid-level descriptors approximating how designers conceptualize lighting, that can be retargeted to arbitrary rigs and edited at the intention level. The focus is on Musical Staging, artistic interpretation that responds to music, rather than functional illumination for visibility.

The proposed Dual Abstraction Layer decomposes lighting behavior into two complementary representations: Global Intentions capture slow-varying visual dynamics across entire songs, while Segment-Based Oscillator patches parameterize rhythmic patterns within musically coherent segments. This separation allows models to specialize: a Diffusion model for continuous global intentions, a Conformer for discrete oscillator patches. An interpreter retargets the outputs to rig-specific DMX and exposes them as an editable interface, ensuring the human designer remains the final author.

The main contributions are: (i) a dual abstraction layer that yields an interpretable, compact parameter space aligned with lighting practice, (ii) an end-to-end generative pipeline combining Diffusion and Conformer models trained on 1,400 minutes of professional show data, (iii) a mixed-method evaluation using intention-based metrics (Overall Quality Score=76.4%) and thematic analysis from an expert study, (iv) empirical validation of rig-retargetability through cross-rig fidelity analysis, and (v) an open-source evaluation pipeline [19].

2 Related Work

This work builds on three research threads: automatic stage lighting control, cross-modal sequence generation, and co-creative systems.

2.1 Automatic Stage Lighting Control

Research on automated stage lighting spans rule-based audio reactivity and data-driven generative approaches. Early ASLC systems emphasize heuristic mappings from spectral features to presets [14]. Hils [5] trains models that map

Table 1. Comparison of ASLC approaches.

Approach	Abstraction	Retargeting	Editing	Evaluation
Hils [5]	Raw params	Limited*	No	Quantitative
Skip-BART [20]	HSV/frame	Limited*	No	Human eval
This work	Dual (Int.+Osc.)	Yes	Yes (UI)	Mixed-method

*Limited retargeting indicates basic color-value transfer without semantic group mapping. Prior systems output 3 parameters (RGB/HSV); this work outputs 72+ parameters across semantic groups.

spectrograms to RGB values for individual fixtures. Kohl et al. [9] propose an intention-based evaluation framework with cross-modal metrics. Zhao et al. [20] propose Skip-BART, which outputs per-frame HSV-like (Hue, Saturation, Value) descriptions for small fixture arrays. These systems operate at fixture parameter level with limited support for rig retargeting and semantic editing. Key differences are presented in Table 1.

Direct comparison with prior systems is infeasible: output representations differ fundamentally (Skip-BART [20] and Hils [5] produce 3-parameter fixture-level outputs vs. the 72+ semantic-group parameters here), and neither codebase nor trained model is publicly available. The evaluation therefore uses internal ablations (Section 4.4) and a naive RMS baseline to isolate architectural contributions.

To our knowledge, no prior ASLC system combines semantic-level retargetability with an editable co-creative interface.

2.2 Cross-Modal Sequence Generation

The design draws on audio-driven motion synthesis. Systems such as EDGE [17] and Bailando [16] use Diffusion and Transformer architectures for dance generation. Alexanderson et al. [1] propose a Conformer-based Diffusion model improving rhythmic alignment. A recurring principle is to operate in a mid-level space rather than predicting raw coordinates directly. These approaches target continuous skeletal motion with fixed output dimensionality, whereas lighting control requires variable rig configurations and discrete rhythmic mechanisms, motivating the dual abstraction presented here.

2.3 Co-Creative Systems

The system is designed as an assistive tool with shared agency and iterative refinement [15]. Following the COFI (Co-Creative Framework for Interaction) framework, it operates in a mixed-initiative mode where the AI proposes lighting abstractions while the designer retains evaluative authority and can override any parameter. Jordanous [6] argues for multidimensional evaluation of creative systems, combining process, product, and person perspectives. This work follows that approach: intention-based metrics assess output quality, while the expert study provides insight into workflow integration and user perception.

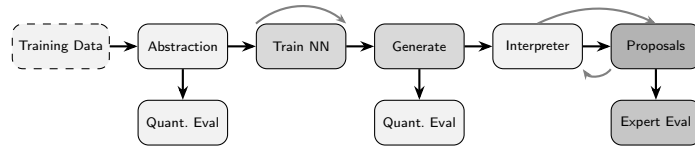


Fig. 1. Generative pipeline. Training data is abstracted and used to train neural network models. Generated outputs pass through an interpreter that enables co-creative editing and retargeting to specific rigs.

Table 2. Dataset comparison.

Dataset	Modality	Duration (min)	Sequences
This Work	Lighting Control (Sync. Audio)	1,400	280
AIST++ [12]	3D Dance Motion	312	1,408
FineDance [11]	3D Dance Motion	876	346

3 System Design

Synchronized audio and lighting are abstracted into global intention trajectories and segment-based oscillator patches. Two generative models operate on this representation, and an interpreter retargets the resulting abstractions to rig-specific DMX-512 control. Figure 1 summarizes the pipeline.

3.1 Dataset and Semantic Grouping

Training requires a corpus of professionally authored synchronized music and lighting. A dataset of approximately 1,400 minutes was assembled, comprising 280 songs from live performances at festivals and touring productions. Audio was aligned with lighting control streams at frame-level precision and standardized to 30 Hz. Train/validation/test splits (80/10/10) were stratified by artist and venue to prevent leakage of show-specific characteristics.

Raw DMX-512 data is mapped to fixtures using per-fixture metadata (modes, channel layouts) and show-specific patch information. Fixtures are grouped into semantic roles based on type, position, and use (e.g., washes, spots/beams, back-line, audience effects). Group taxonomies are defined per-show using patch metadata, and the interpreter maps these to target rigs via explicit user assignment. Grouping reduces dimensionality and forms the basis for a rig-agnostic intention space. Cross-rig mapping fidelity is evaluated in Section 4.3. Table 2 contextualizes the corpus scale against benchmarks from related domains.

3.2 Dual Abstraction Layer

Professional lighting design operates at multiple temporal scales: broad emotional arcs across a song (density, brightness, color following musical form) and

tight rhythmic patterns within sections (choreographic associations to beats and onsets). Early experiments showed that a single abstraction cannot adequately capture both scales.

The Dual Abstraction Layer therefore separates (i) a continuous global intention stream at frame level, capturing slow-varying aesthetic dynamics, and (ii) a discrete segment-based oscillator stream per musical segment, capturing rhythmic mechanisms. Both act on semantic groups but capture complementary aspects, enabling the system to generate smoothly evolving scenes and sharply synchronized effects.

3.3 Global Intention Representation

The Global Intention representation defines a lighting show as continuous feature curves for each semantic group. The term “intention” here denotes statistical proxies for aesthetic behavior (intensity envelopes, color trajectories) rather than high-level semantic concepts; the representation captures patterns correlated with artistic decisions without modeling designer reasoning directly. For each group, a 6D feature vector is computed per frame, yielding a 72D intention vector ($12 \text{ groups} \times 6 \text{ features}$). All values are normalized to $[0, 1]$. The features capture intensity dynamics (I_{peak} : peak brightness, $\nabla_s I$: spatial gradient distinguishing sharp accents from smooth washes, ρ_{peak} : density of intensity peaks indicating spatial complexity, $I_{\text{min_inv}}$: contrast proxy), and color ($\bar{H}, \bar{S}$: mean hue and saturation).

Figure 2 illustrates how chaotic individual fixture signals are transformed into smooth, low-dimensional, semantically rich curves suitable for learning. A Diffusion model generates global intention trajectories conditioned on audio features, following the EDGE architecture [17].

3.4 Segment-Based Oscillator Representation

Global intentions capture broad dynamics but do not explicitly parameterize rhythmic mechanisms. Analogous to audio synthesis, this layer represents lighting patterns within segments as parametric oscillator patches, changing the learning task from predicting continuous signals to predicting compact, editable parameters.

Within a segment, designers often employ recognizable patterns such as square-wave strobes, sine-like pulses, and ramps. For segment s and group g , intensity is approximated by $V_{g,s}(t) = A \cdot \text{Wave}(\omega t + \phi) + O$, where $\text{Wave} \in \{\text{sine, square, sawtooth, triangle, gaussian, static}\}$, ω is frequency (normalized to $[0, 8] \text{ Hz}$), ϕ is phase (normalized to $[0, 2\pi]$), and $A, O \in [0, 1]$ are amplitude and baseline offset. Each patch also includes mean hue/saturation and a Movement Activity Index (MAI) summarizing fixture pan/tilt motion intensity over the segment. Segments are derived from audio using the All-in-One model [7].

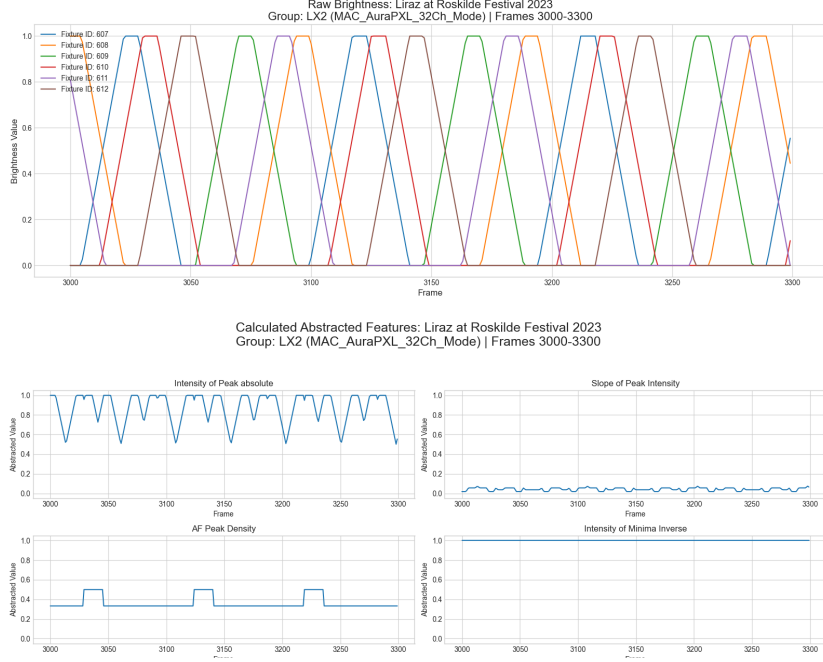


Fig. 2. Abstraction efficacy. Top: raw brightness channels for a fixture group. Bottom: corresponding global intention features showing smooth, interpretable curves.

3.5 Generative Models and Conditioning

Both models are conditioned on audio features at 30 Hz: onset curves and beat positions (madmom [3]), chroma-based self-similarity, and jukebox embeddings [4]. The Global Intention Diffusion model ($\sim 49\text{M}$ parameters) uses 8 Conformer blocks following EDGE [17], trained for 200 epochs. The Segment-Based Oscillator model uses a Conformer following Alexanderson et al. [1], trained for 150 epochs.

3.6 Interpreter and Co-Creative Editing

The interpreter maps intentions to a target rig by assigning semantic groups to fixtures and rendering oscillator patches into control trajectories. During back-conversion, the continuous Global Intention stream provides the atmospheric envelope while segment-level oscillator patches supply local rhythmic detail. Both layers combine multiplicatively per group through a hand-tuned weighting matrix W , yielding $V_{\text{final}}(t) = W \odot V_{\text{GI}}(t) \odot V_{\text{Osc}}(t)$. At segment boundaries derived from the audio analysis, the active oscillator patch updates without crossfading, and each user edit triggers recomputation for immediate preview. The interpreter enforces safety constraints and exposes controllable parameters through

a combined WebUI and physical knob/fader controller, providing real-time visual feedback from the visualizer or physical rig. This ensures the human designer remains the final author. The system targets offline pre-production with real-time playback of pre-generated abstractions. Retargetability is validated in Section 4.3.

4 Evaluation

Evaluating generative lighting requires attention to both measurable correspondences and subjective perception. This work adopts a mixed-method design combining intention-based metrics with qualitative expert feedback [6, 10].

4.1 Evaluation Setup

Two datasets support the evaluation: (i) the professional corpus providing the reference distribution, and (ii) a set of 51 tracks unseen during model training, used as audio-only inputs. For each metric M , the Achievement Ratio is defined as $AR_M = \text{median}(M_{\text{gen}}) / \text{median}(M_{\text{gt}})$, comparing generated outputs to professional references. Values above 1.0 indicate that the system produces more of the measured behavior than the professional reference, and whether this is desirable depends on the metric.

4.2 Quantitative Results

Self-Similarity Matrix (SSM) correlation measures how well lighting changes reflect musical structure. The system achieves SSM correlation of 0.162 versus 0.238 for professional references (AR 68.1%), indicating moderate structural correspondence. Functional novelty correlation measures whether major lighting transitions align with musical boundaries such as verse-chorus changes. The system reaches AR 82.2%, confirming that transitions are placed near musically salient moments.

Beat peak alignment (AR 125.7%) and onset change correlation (AR 99.3%) indicate strong rhythmic responsiveness, though alignment above the professional reference should not be read as unambiguously superior: professionals routinely pull back from strict beat alignment for expressive effect.

The Overall Quality Score (OQS) serves as a development benchmark, comparing generated outputs to professional references as a weighted composite (structural/temporal 0.16, intention-based comparison 0.42, oscillator evaluation 0.42), weighting the two areas that benchmark directly against professional ground truth more heavily than internal musical alignment, with each area capped at 100%. The system achieves an OQS of 76.4%, combining structural/temporal analysis (88.5%), intention-based reference comparison (80.4%), and oscillator evaluation (67.9%). The lower oscillator sub-score reflects the difficulty of reconstructing reference-like patch shapes, since wave-type classification admits multiple musically valid solutions, even though the oscillator layer

Table 3. Intention Preservation Under Cross-Rig Application. Pearson correlations measuring temporal dynamics preservation.

Rig Configuration	Positions	Mean I	Peak I	Range	Hue
Direct Mapping [†]	8	0.40	0.28	0.45	0.71
Club Rig	8*	0.80	0.62	0.87	0.78
Concert Rig	19	0.82	0.75	1.00	0.86
LED Bar Array	21	0.71	0.79	1.00	0.87
Reference (Ours)	33	1.00	1.00	1.00	1.00

[†]Direct Mapping uses naive truncation without spatial awareness. *Center-mirrored.

is central to localized rhythmic detail (evaluation framework available as open source [18]).

4.3 Cross-Rig Retargeting Validation

To validate the retargetability claim, generated proposals were adapted to three target rig configurations with varying spatial resolutions: a club rig (16 moving heads, 8 unique positions with center-mirroring), a concert rig (56 fixtures, 19 unique positions), and an LED bar array (64 pixels, 21 unique positions). A naive direct mapping baseline uses truncation without spatial awareness. Table 3 reports temporal correlation metrics comparing adapted outputs to the native 33-position reference. Mean I and Peak I measure temporal intensity dynamics preservation, Range captures dynamic contrast fidelity, and Hue measures color consistency. The evaluation code is available in the companion repository [19].

At the same spatial resolution (8 positions), intention-preserving resampling achieves $2\times$ better mean intensity correlation than naive mapping (0.80 vs. 0.40), and concert-scale rigs (19+ positions) achieve near-perfect dynamic range preservation ($r=1.00$).

4.4 Ablation: Dual-Model Contribution

In the absence of directly comparable prior systems (Section 2.1), ablation analysis isolates the contribution of each architectural component. Three conditions were evaluated on the same validation tracks: the full dual-model system, Diffusion-only generation (oscillator branch disabled, static intensity within segments), and oscillator-only generation (oscillator patterns rendered over a constant intensity envelope). An RMS baseline maps audio energy directly to brightness without learned patterns, representing naive audio-reactive behavior. Table 4 reports results normalized to the full system (ablation evaluation code available in the companion repository [19]).

Diffusion-only generation retains 77% of structural correspondence but only 43% of beat alignment, confirming that the Diffusion model captures broad dynamics while lacking rhythmic precision. Oscillator-only generation achieves 137% beat peak alignment but degrades to 62% on structure, demonstrating

Table 4. Ablation Study (normalized to Full=100%). RMS (Root Mean Square) Baseline uses naive audio energy mapping without learned patterns.

Condition	SSM	Corr	Novelty	Beat	Peak	Beat	Valley
Full (Ours)	100%	100%	100%	100%	100%	100%	100%
Diffusion-only	77%	57%	43%	64%			
Oscillator-only	62%	79%	137%	136%			
RMS Baseline	59%	47%	90%	98%			



Fig. 3. Experimental setup at the Sound and Immersive Media Laboratory, HAW Hamburg, Faculty of Electrical Engineering, Media and Information Technology.

complementary specialization. The oscillator branch’s superior rhythmic precision when isolated reflects its design for beat-aligned patterns; however, this precision comes at the cost of song-level coherence. The full system’s AR of 125.7% suggests the diffusion layer moderates but does not fully suppress this tendency. The full system therefore trades peak rhythmic alignment for more balanced performance across both temporal scales. The RMS baseline achieves only 59% structural correspondence and 90% beat alignment, confirming that learned patterns substantially outperform naive audio-reactive mapping on structural metrics while providing comparable rhythmic response.

4.5 Exploratory Expert Study

An exploratory study was conducted with two professional lighting designers ($N=2$), each with more than 10 years of experience. Participants generated proposals for unseen tracks, edited parameters, and observed results on a visualizer and physical rig. Sessions were analyzed using thematic analysis; at $N=2$ the resulting themes indicate workflow value and failure modes rather than validating quantitative claims.

Three themes emerged:

Theme 1 (Musician-like timing): Participants described the timing as “musician-like,” highlighting syncopations and rhythmic detail that would be laborious to program manually, consistent with high beat/onset alignment metrics. Whether this responsiveness remains desirable across full-length performances is an open question given this sample size.

Theme 2 (Ecosystem integration): The main barrier to adoption was the standalone nature of the prototype. Integration into industry-standard software ecosystems and workflows was identified as essential for production use.

Theme 3 (Always-on trade-off): Generated shows tended to avoid prolonged blackouts, reducing negative space and dynamic contrast.

5 Discussion and Limitations

The evaluation shows that the dual abstraction yields musically responsive and editable proposals, but the rhythmic results warrant careful interpretation. Beat alignment above the professional reference may reflect over-reactivity rather than superior musical judgment: experienced designers routinely delay or withhold cue responses for expressive effect, and the current system has no mechanism for such intentional restraint. The expert study’s positive reception of timing quality (Theme 1) is consistent with both interpretations at $N = 2$. Explicit modeling of beat restraint, for instance through a learned attenuation parameter, is a direction for future work. Cross-rig validation (Table 3) confirms that intention-preserving resampling achieves $2\times$ better temporal correlation than naive mapping at the same spatial resolution, a capability absent from prior generative approaches that bind outputs to specific rigs. The ablation study justifies the architectural separation of continuous dynamics from segment-level rhythmic mechanisms.

Genre scope and generalizability. Trained primarily on rhythmic concert data, the system’s Segment-Based Oscillator assumes periodic beat alignment. This design restricts applicability to non-rhythmic contexts (e.g., theatre, free jazz), where generalization remains untested.

Long-range structure. SSM correlation remains moderate, suggesting that motif reuse and song-level narrative are not consistently captured. Hierarchical planning or memory-augmented architectures could improve structural coherence.

Negative space. The “always-on” behavior ensures designers always have an editable proposal, but limits dynamic contrast. An explicit blackout state is a direction for future work.

Ecosystem integration. Native integration into industry-standard ecosystems (e.g., grandMA [13] plugins, Art-Net/sACN bridges) remains essential for deployment.

Segmentation dependence. Segment boundaries from an external model [7] can propagate errors to oscillator prediction.

Expert study scope. The qualitative evaluation is preliminary ($N=2$). Recruiting designers with the required expertise for hands-on sessions remains challenging; larger-scale validation and human preference studies for cross-rig outputs remain future work.

Reproducibility. The full corpus cannot be released due to NDAs; however, a subset of abstracted training data from the Roskilde Festival is available [8], alongside the abstraction and evaluation pipeline with example data [19].

6 Conclusion

This paper presented an end-to-end system for music-driven stage lighting based on a dual abstraction layer. By representing lighting in terms of global intentions and segment-based oscillator patches, the system separates structural dynamics from rhythmic mechanisms and exposes an editable parameter space. A Diffusion model and Conformer generate these abstractions from audio, and an interpreter retargets them to rig-specific control.

Mixed-method evaluation demonstrates an Overall Quality Score of 76.4% relative to professional references, with cross-rig validation confirming 2× better intention preservation than naive mapping. Strengths lie in rhythmic responsiveness, though the system does not yet model deliberate restraint; limitations remain in long-range structure and negative space. The results support abstraction-centered generation as a co-creative starting point that augments designer expertise while preserving human agency. Beyond stage lighting, the success of the dual abstraction suggests a transferable approach for other creative domains where a large semantic gap exists between raw technical parameters and high-level artistic intentions.

Acknowledgments. This work was supported by HAW Hamburg. The authors thank Sebastian Jakob for professional insight into lighting design practice, Prof. Dr. Roland Greule for encouraging this research, and the professional lighting designers who participated in the evaluation study. MA Lighting Technology GmbH supported this work through resources and expertise.

References

1. Alexanderson, S., Nagy, R., Beskow, J., Henter, G.E.: Listen, denoise, action! audio-driven motion synthesis with diffusion models. *ACM Transactions on Graphics* **42**(4), 1–20 (2023). <https://doi.org/10.1145/3592458>
2. ANSI: ANSI E1.11 – 2008 (R2018): Entertainment technology – USITT DMX512-A – asynchronous serial digital data transmission standard for controlling lighting equipment and accessories. Tech. rep., Entertainment Services and Technology Association (2008), reaffirmed 2018
3. Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F., Widmer, G.: madmom: A new python audio and music signal processing library. In: *Proceedings of the 24th ACM International Conference on Multimedia*. pp. 1174–1178 (2016). <https://doi.org/10.1145/2964284.2973795>
4. Dhariwal, P., Jun, H., Payne, C., Kim, J.W., Radford, A., Sutskever, I.: Jukebox: A generative model for music (2020). <https://doi.org/10.48550/ARXIV.2005.00341>, <https://arxiv.org/abs/2005.00341>
5. Hils, M.: Deep Learning for Sound-to-Light Automation in Stage Lighting Applications. M.s. thesis, Technical University of Munich (2023), https://collab.dvb.bayern/download/attachments/191204145/Deep_Learning_for_Sound_to_Light_Automation_in_Stage_Lighting_Applications.pdf
6. Jordanous, A.: A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation* **4**(3), 246–279 (2012). <https://doi.org/10.1007/s12559-012-9156-1>

7. Kim, T., Nam, J.: All-in-one metrical and functional structure analysis with neighborhood attentions on demixed audio. In: 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). pp. 1–5. IEEE (2023). <https://doi.org/10.48550/ARXIV.2307.16425>, <https://arxiv.org/abs/2307.16425>
8. Kohl, M.: Cross-modal metrics for capturing correspondences in stage performances: Dataset and evaluation code. <https://github.com/MKKeys92/Cross-Modal-Metrics-for-Capturing-Correspondences-in-Stage-Performances> (2025), gitHub repository with Roskilde Festival dataset. Accessed: 2026-01-14
9. Kohl, M., Wursthorn, T., Weiß, C.: Cross-modal metrics for capturing correspondences between music audio and stage lighting signals. In: Proceedings of the 33rd ACM International Conference on Multimedia (MM '25). pp. 528–534. ACM (2025). <https://doi.org/10.1145/3746027.3755488>
10. Lazar, J., Feng, J.H., Hochheiser, H.: Research Methods in Human-Computer Interaction. Morgan Kaufmann, 2nd edn. (2017)
11. Li, R., Zhao, J., Zhang, Y., Su, M., Ren, Z., Zhang, H., Tang, Y., Li, X.: Finedance: A fine-grained choreography dataset for 3D full body dance generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10234–10243 (2023). <https://doi.org/10.1109/ICCV51070.2023.00939>
12. Li, R., Yang, S., Ross, D.A., Kanazawa, A.: AI choreographer: Music conditioned 3D dance generation with AIST++. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13401–13412 (2021). <https://doi.org/10.1109/ICCV48922.2021.01315>
13. MA Lighting: MA Lighting – professional lighting control. <https://www.malighting.com/> (2025), accessed: 2026-01-06
14. Oxholm, E., Hansen, E.K., Triantafyllidis, G.: Auditory and visual based intelligent lighting design for music concerts. EAI Endorsed Transactions on Creative Technologies **5**(15), e2 (2018). <https://doi.org/10.4108/eai.10-4-2018.154452>
15. Rezwana, J., Maher, M.L.: Designing creative AI partners with CoFI: A framework for modeling interaction in human-AI co-creative systems. ACM Transactions on Computer-Human Interaction **30**(5) (2023). <https://doi.org/10.1145/3519026>
16. Siyao, L., Yu, W., Gu, T., Lin, C., Wang, Q., Qian, C., Loy, C.C., Liu, Z.: Bailando: 3D dance generation by actor-critic GPT with choreographic memory. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1171–1181 (2022). <https://doi.org/10.1109/CVPR52688.2022.00124>
17. Tseng, J., Castellon, R., Liu, C.K.: EDGE: Editable dance generation from music (2022). <https://doi.org/10.48550/ARXIV.2211.10658>, <https://arxiv.org/abs/2211.10658>
18. Wursthorn, T.: Cross-modal metric evaluation framework for music-driven stage lighting. https://github.com/twotonetobi/MSc_thesis_metric_evaluation (2026), gitHub repository. Accessed: 2026-01-14
19. Wursthorn, T.: End-to-end music-driven stage lighting: Implementation and evaluation pipeline. https://github.com/twotonetobi/End-to-End_Music-Driven_Stage_Lighting_2026 (2026), gitHub repository. Accessed: 2026-01-06
20. Zhao, Z., Jin, D., Zhou, Z., Zhang, X.: Automatic stage lighting control: Is it a rule-driven process or generative task? (2025). <https://doi.org/10.48550/ARXIV.2506.01482>, <https://arxiv.org/abs/2506.01482>