Synesthetic Algorithmic Composition: A Bio-Inspired Framework for Cross-Modal Creative Generation

Dr. Elara Vance* Prof. Kenji Tanaka† Dr. Isabella Rossi‡

Introduction

The intersection of computational creativity and multi-modal perception represents a frontier in artificial intelligence research with profound implications for both artistic practice and cognitive science. While significant advances have been made in generative systems for individual creative domains, the computational modeling of cross-modal creative expression remains largely unexplored. This paper addresses this gap by introducing a novel framework for synesthetic algorithmic composition that generates intrinsically linked auditory and visual artworks through bio-inspired neural mechanisms.

Synesthesia, the neurological phenomenon where stimulation of one sensory pathway leads to automatic experiences in a second pathway, provides a compelling model for computational creativity. Previous approaches to multi-modal generation have typically employed separate systems for different modalities with post-hoc alignment, resulting in outputs that lack the deep semantic integration characteristic of human cross-modal perception. Our work fundamentally rethinks this paradigm by developing computational models that mirror the neural underpinnings of synesthetic experience.

We propose three research questions: (1) How can we computationally model the structural mappings between auditory and visual creative domains? (2) What architectural principles enable the generation of coherent multi-modal artistic outputs? (3) To what extent do such computational models capture meaningful aspects of human cross-modal perception?

Our contributions include: the Cross-Modal Resonance Network architecture, a novel dataset of aligned musical and visual artworks, and empirical validation through both computational metrics and human evaluation. This research

*Department of Computational Creativity, University of Cambridge
†Media Arts and Technology, Tokyo Institute of Technology
‡Cognitive Science Laboratory, University of Milan

advances computational creativity while providing new tools for exploring the nature of cross-modal artistic expression.

Methodology

Architectural Framework

Our Cross-Modal Resonance Network (CMRN) consists of three interconnected components designed to model different aspects of cross-modal creative generation:

Hierarchical Temporal Encoder (HTE): This component processes musical inputs across multiple temporal scales using a combination of convolutional and recurrent neural networks. The encoder captures features from micro-level musical events (individual notes and chords) to macro-level structural patterns (phrases, sections, and overall form).

Spatial-Semantic Mapping Engine (SSME): Inspired by the neural organization of sensory cortices, this module establishes bidirectional mappings between auditory features and visual representations. The mapping employs a novel attention mechanism that learns to associate musical patterns with visual elements based on both structural and emotional characteristics.

Bidirectional Generative Interface (BGI): This component enables creative generation in both directions (music-to-visual and visual-to-music) through a variational autoencoder framework with cross-modal constraints. The interface ensures that generated outputs maintain coherence within their modality while exhibiting meaningful relationships across modalities.

Mathematical Formulation

Let M represent musical space and V represent visual space. Our framework learns a mapping function $\Phi : M \leftrightarrow V$ such that for any musical sequence $m \in M$ and visual artwork $v \in V$:

$$\Phi(m, v) = \arg\min_{\theta} \left[\mathcal{L}_{recon}(m, v) + \lambda \mathcal{L}_{cross}(m, v)\right] \tag{1}$$

where $\mathcal{L}_{recon}$ ensures faithful reconstruction within each modality, and $\mathcal{L}_{cross}$ enforces cross-modal coherence through a novel similarity metric that combines structural, emotional, and semantic alignment.

Training and Implementation

We trained our system on a curated dataset of 15,000 paired musical compositions and visual artworks spanning classical, jazz, electronic, and contemporary

genres. The training employed a multi-stage approach with progressive complexity, beginning with simple melodic and color relationships and advancing to complex structural and emotional mappings.

Results

Quantitative Evaluation

We evaluated our framework against three baseline approaches: separate modality systems with post-hoc alignment, traditional multi-modal VAEs, and transformer-based cross-modal models. Evaluation metrics included cross-modal coherence, within-modality quality, and human perception scores.

Table 1: Performance Comparison Across Methods

Method	Cross-Modal Coherence	Musical Quality	Visual Quality
CMRN (Ours)	**4.2/5.0**	4.5/5.0	4.3/5.0
Separate Systems	2.8/5.0	4.6/5.0	4.4/5.0
Multi-modal VAE	3.1/5.0	4.2/5.0	4.1/5.0
Transformer	3.4/5.0	4.3/5.0	4.2/5.0

Our approach demonstrated superior performance in cross-modal coherence while maintaining competitive performance within individual modalities. The 37

Qualitative Analysis

Human evaluators consistently noted that CMRN-generated pairs exhibited deeper semantic relationships than baseline approaches. For example, musical passages with ascending melodic lines frequently generated visual compositions with upward-moving elements, while rhythmically complex sections produced visually dense patterns. These emergent mappings aligned with established psychological theories of cross-modal perception.

Emergent Patterns

Analysis of the learned mappings revealed several intriguing patterns:

- High-frequency musical elements consistently mapped to bright, high-contrast visual regions
- Harmonic complexity correlated with visual texture density
- Emotional valence in music showed strong correspondence with color saturation in visuals

- Rhythmic patterns influenced spatial repetition and symmetry in visual outputs

These patterns suggest that our computational model captures meaningful aspects of human cross-modal perception beyond superficial feature correlations.

Conclusion

This paper has introduced a novel framework for synesthetic algorithmic composition that bridges auditory and visual creative domains through bio-inspired neural mechanisms. Our Cross-Modal Resonance Network architecture represents a significant departure from traditional multi-modal generation approaches by modeling the intrinsic relationships between sensory modalities rather than treating them as separate systems.

The experimental results demonstrate that our approach generates multi-modal artistic outputs with substantially higher cross-modal coherence than existing methods, while maintaining high quality within individual modalities. The emergent patterns in the learned mappings provide computational evidence for theories of cross-modal perception and suggest new directions for both artistic practice and cognitive science research.

Future work will explore extensions to additional sensory modalities, real-time interactive applications, and deeper integration with neurological models of synesthesia. The framework also has potential applications in therapeutic environments, educational tools, and enhanced creative interfaces for artists working across sensory boundaries.

Our research contributes to the growing field of computational creativity while providing new insights into the fundamental nature of cross-modal artistic expression. By modeling the neurological phenomenon of synesthesia, we have developed computational tools that not only generate compelling artistic outputs but also help illuminate the complex relationships between different forms of human creative expression.

References

1. Cytowic, R. E. (2002). Synesthesia: A Union of the Senses. MIT Press.
2. Ward, J. (2013). Synesthesia. Annual Review of Psychology.
3. Manovich, L. (2001). The Language of New Media. MIT Press.
4. Zatorre, R. J., et al. (2007). Structure and function of auditory cortex. Nature Reviews Neuroscience.
5. Goodfellow, I., et al. (2014). Generative Adversarial Networks. NeurIPS.

6. Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. ICLR.