Synesthetic Encoding: A Bio-Inspired Framework for Cross-Modal Data Representation in Neuromorphic Systems

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

Introduction

The human brain possesses remarkable capabilities for integrating information across sensory modalities, a phenomenon most dramatically illustrated in synesthesia, where stimulation of one sensory pathway leads to automatic experiences in another. While artificial intelligence has made significant strides in processing individual sensory modalities, current systems largely lack the rich cross-modal integration that characterizes human perception. This paper introduces Synesthetic Encoding, a computational framework that bridges this gap by implementing bio-inspired cross-modal associations in neuromorphic architectures.

Traditional multimodal approaches typically process sensory streams in parallel and combine them at later stages, missing the fundamental integration that occurs in biological systems. Our work is distinguished by its direct modeling of cross-modal neural pathways, creating a system where representations in one modality inherently contain information about other modalities. This approach represents a paradigm shift from conventional multimodal processing toward truly integrated sensory representation.

Our research addresses three fundamental questions: (1) How can we computationally model the neural mechanisms underlying synesthetic experiences? (2) What architectural principles enable efficient cross-modal information transfer in artificial neural systems? (3) What practical advantages do synesthetic representations offer over traditional multimodal approaches in real-world applications?

[*]Neurocomputing Research Lab, University of Cambridge
[†]Institute for Advanced AI Studies, Kyoto University
[‡]Creative Computing Department, Politecnico di Milano

Methodology

Neuromorphic Architecture

We designed a hierarchical spiking neural network with three primary sensory encoders for visual, auditory, and tactile data. Each encoder transforms raw sensory input into sparse temporal spike patterns using leaky integrate-and-fire neurons. The core innovation lies in the Cross-Modal Projection Layers (CMPL), which establish bidirectional connections between sensory domains through learned weight matrices.

The network architecture follows:

$$S_{target} = f(W_{source \rightarrow target} \cdot S_{source} + b) \tag{1}$$

Where S_{source} and S_{target} represent spike patterns in source and target modalities, W denotes learned cross-modal projection weights, and b represents bias terms.

Neural Resonance Training

We developed a novel training algorithm called Neural Resonance Training (NRT) that simultaneously optimizes unimodal representations and cross-modal projections. Unlike conventional backpropagation, NRT employs a biologically-plausible spike-timing-dependent plasticity rule combined with a resonance objective function that maximizes mutual information between associated cross-modal patterns.

The resonance objective is defined as:

$$\mathcal{L}_{resonance} = \sum_{i,j} I(S_i; S_j) - \lambda D_{KL}(P_{cross} || P_{independent}) \tag{2}$$

Where $I(S_i; S_j)$ represents mutual information between spike patterns in modalities i and j, and the KL-divergence term regularizes the cross-modal associations.

Experimental Design

We evaluated our framework on three distinct applications:

1. **Cross-modal retrieval**: Assessing the system's ability to retrieve images given audio queries and vice versa using the Multimodal Association Dataset (MAD-200k)

2. **Sensory substitution**: Testing visual-to-auditory translation for accessibility applications with 15 visually impaired participants

3. **Creative generation**: Exploring emergent cross-modal creativity through music-to-visual-art translation

Comparative baselines included multimodal transformers, cross-modal autoencoders, and traditional feature concatenation approaches.

Results

Cross-Modal Retrieval Performance

Our synesthetic encoding framework demonstrated superior performance in cross-modal retrieval tasks, achieving a mean average precision of 0.87 compared to 0.59 for the best-performing baseline (multimodal transformer). The system particularly excelled in abstract concept retrieval, where semantic relationships between modalities were complex and nuanced.

Table 1: Cross-Modal Retrieval Performance (mAP)

Method	Image→Audio	Audio→Image	Text→Image
Synesthetic Encoding	0.85	0.89	0.87
Multimodal Transformer	0.57	0.61	0.59
Cross-Modal Autoencoder	0.48	0.52	0.50
Feature Concatenation	0.41	0.45	0.43

Computational Efficiency

The neuromorphic implementation of synesthetic encoding demonstrated significant efficiency gains, processing cross-modal queries with 63

Sensory Substitution Applications

In accessibility testing, visually impaired participants achieved 89

Emergent Creative Generation

Perhaps the most surprising result emerged in creative applications. When translating musical compositions into visual artworks, the system produced aesthetically coherent images that preserved the emotional tone and structural patterns of the original music. Independent art critics rated the generated artworks as 'expressive' and 'conceptually integrated' with the source music in 78

Figure 1: Example of cross-modal generation: Beethoven's Moonlight Sonata translated to visual artwork through synesthetic encoding

Conclusion

This paper has introduced Synesthetic Encoding as a novel framework for cross-modal data representation that draws inspiration from human synesthesia. Our approach represents a significant departure from conventional multimodal processing by establishing direct, learned mappings between sensory domains within a neuromorphic architecture.

The key contributions of this work include:

1. A bio-inspired computational model of cross-modal perception based on principles of neural synesthesia 2. The Neural Resonance Training algorithm for simultaneously optimizing unimodal and cross-modal representations 3. Demonstration of practical advantages in retrieval efficiency, accessibility applications, and emergent creative capabilities 4. Evidence that integrated cross-modal representations can surpass traditional approaches in both performance and efficiency

Our findings suggest that moving beyond isolated sensory processing toward more integrated, human-like representations offers substantial benefits for artificial perception systems. The spontaneous emergence of creative cross-modal translations points to the rich representational capacity of synesthetic encoding and opens new possibilities for artificial systems that more closely mimic human cognitive processes.

Future work will explore extensions to additional sensory modalities, applications in human-computer interaction, and investigations into whether similar cross-modal representations can enhance general artificial intelligence capabilities beyond specialized sensory tasks.