

Synesthetic Computing: A Multi-Modal Neural Architecture for Cross-Sensory Data Representation and Processing

Dr. Elara Vance* Prof. Kenji Tanaka[†] Dr. Sofia Petrova[‡]

Abstract

This paper introduces Synesthetic Computing, a novel computational paradigm inspired by neurological synesthesia that enables cross-modal data representation and processing. Unlike traditional unimodal approaches, our framework allows data from one sensory domain (e.g., auditory) to be automatically represented and processed in another domain (e.g., visual or tactile). We developed a multi-layer neural architecture that learns cross-modal mappings through a combination of transformer-based attention mechanisms and generative adversarial networks. Experimental results demonstrate that our approach achieves 89.3% accuracy in cross-modal classification tasks and enables novel applications in accessibility technology, creative computing, and data visualization. The system successfully translates musical compositions into visual art while preserving emotional content, converts textual descriptions into tactile patterns, and transforms visual scenes into auditory experiences with 76.8% semantic preservation according to human evaluators.

Introduction

Traditional computing systems process data within isolated sensory modalities, mirroring conventional human-computer interaction paradigms. However, this approach fails to capture the rich cross-modal associations that occur naturally in human perception and neurological conditions like synesthesia. We propose Synesthetic Computing as a radical departure from modality-specific processing, enabling computers to represent and process information across sensory boundaries.

Our research addresses three fundamental questions: (1) Can computational systems learn meaningful cross-modal mappings without explicit supervision?

*Department of Computational Neuroscience, Stanford University

[†]Media Arts and Technology, University of Tokyo

[‡]Creative Computing Lab, MIT Media Lab

(2) How can we quantify the preservation of semantic content across sensory transformations? (3) What novel applications emerge when computers can "experience" data through multiple sensory channels simultaneously?

The novelty of our approach lies in its biological inspiration and architectural design. Rather than treating cross-modal translation as a simple conversion task, we model it as a holistic representation learning problem where the system develops an internal "sensory vocabulary" that transcends individual modalities.

Methodology

Architectural Overview

Our Synesthetic Neural Architecture (SNA) consists of three core components:

- **Modality-Specific Encoders:** Separate transformer-based networks that extract hierarchical features from visual, auditory, textual, and tactile inputs
- **Cross-Modal Attention Bridge:** A novel attention mechanism that learns alignment between different sensory representations
- **Multi-Modal Generator:** A conditional GAN architecture that synthesizes outputs in target modalities while preserving semantic content

Mathematical Formulation

Let $M = \{V, A, T, H\}$ represent our four primary modalities (Visual, Auditory, Textual, Haptic). For any input $x_i \in M_i$, the system learns a mapping:

$$\Phi : M_i \rightarrow M_j \quad \forall i, j \in M \quad (1)$$

Through a shared latent space Z , where the transformation preserves semantic content measured by:

$$S(\Phi(x_i), x_j) > \tau \quad \text{for semantically equivalent pairs} \quad (2)$$

Training Protocol

We trained the SNA on a novel dataset comprising 50,000 multi-modal examples, including:

- Musical compositions paired with abstract visual art created by the same artist
- Text descriptions of tactile experiences with corresponding haptic feedback patterns

- Environmental sounds with matching visual scenes
- Emotional narratives across different expressive forms

The training employed a multi-objective loss function combining reconstruction loss, semantic preservation loss, and cross-modal consistency loss.

Results

Quantitative Evaluation

Table 1: Cross-Modal Classification Accuracy (%)

Target Modality	Visual	Auditory	Textual	Haptic
Visual	-	87.2	91.5	83.7
Auditory	85.9	-	88.3	79.4
Textual	92.1	89.7	-	86.2
Haptic	81.3	76.8	84.9	-

Qualitative Assessment

Human evaluators (n=150) assessed the quality of cross-modal transformations:

- **Emotional Preservation:** 82% of participants agreed that emotional content was maintained across sensory transformations
- **Semantic Coherence:** 76.8% average agreement on semantic preservation across all modality pairs
- **Aesthetic Quality:** Visual and auditory outputs received particularly high ratings for creative appeal

Novel Applications

We demonstrated three groundbreaking applications:

Synaesthetic Art Generation: The system generated visual artworks from musical inputs that were rated as more emotionally congruent than traditional algorithmic art ($p < 0.01$).

Multi-Sensory Data Exploration: Users could "hear" statistical patterns or "feel" network structures, enabling intuitive understanding of complex datasets.

Accessibility Enhancement: The system created tactile representations of visual information for visually impaired users, with 94% reporting improved comprehension.

Conclusion

Synesthetic Computing represents a paradigm shift in how computers represent and process information. Our results demonstrate that neural systems can learn meaningful cross-modal mappings that preserve semantic and emotional content. The applications span creative computing, accessibility technology, data visualization, and human-computer interaction.

Future work will explore extending the framework to additional sensory modalities (olfactory, gustatory) and developing real-time synesthetic interfaces. The philosophical implications of computers developing cross-sensory experiences warrant further investigation, particularly regarding artificial consciousness and perception.

Our contribution lies not only in the technical architecture but in redefining the boundaries of computational perception. By enabling machines to transcend traditional sensory silos, we open new possibilities for creative expression, accessibility, and human-machine collaboration.