

## 인공감각: 시청각-촉각 상호생성을 통한 실감형 미디어 전송 프레임워크

유동호<sup>1</sup>, 김병현<sup>2</sup>, 이용준<sup>2</sup>, 이승석<sup>2</sup>, 전승엽<sup>2</sup>, 정승혁<sup>2</sup>국립공주대학교<sup>1</sup>, 한남대학교<sup>2</sup>dongho.you@kongju.ac.kr<sup>1</sup>

## Artificial Sensation: A Cross-modal Generation Framework for Haptic-Audiovisual Immersive Media Transmission

Dongho You<sup>1</sup>, Byunghyun Kim<sup>2</sup>, Yongjun Lee<sup>2</sup>, Seungseok Lee<sup>2</sup>, Seungyep Jeon<sup>2</sup>, Seunghyeok Jeong<sup>2</sup>Kongju National University<sup>1</sup>, Hannam University<sup>2</sup>

## 요약

본 논문은 시각·청각·촉각 간 상호생성을 가능하게 하는 새로운 개념인 인공감각(Artificial Sensation) 프레임워크를 제안한다. 제안된 접근은 인공지능이 시청각 정보로부터 촉각을 예측하고, 반대로 촉각 신호로부터 시청각 경험을 복원함으로써 센서 없는 실감형 미디어 전송의 패러다임을 제시한다.

## I. 서론

현대의 실감형 미디어는 주로 시각과 청각을 중심으로 발전해왔다. 가상현실(VR)과 증강현실(AR) 기술 진보에도 불구하고, 촉각은 여전히 한정적인 역할에 머무르며, 대부분의 시스템은 시청각 정보를 보조하기 위한 제한된 형태의 진동 자극이나 피드백 장치로 사용되고 있다<sup>[1-3]</sup>. 이러한 접근은 인간의 실제 감각 구조와는 괴리가 있다. 인간의 지각은 기본적으로 다감각적이며, 감각들은 상호 보완적 관계 속에서 통합되어 하나의 경험을 구성한다. 시각이 차단되면 청각과 촉각이 강화되고, 소리만으로도 물체의 질감과 거리, 무게를 상상할 수 있다. 즉, 감각은 단절된 신호가 아니라 서로의 의미를 추론하고 예측하는 지능적 체계라 할 수 있다.

본 연구는 이러한 인간의 감각의 통합성과 예측성을 인공지능 기반 미디어 구조에 적용하여, 센서 없이도 시청각 정보만으로 촉각을 생성하고, 반대로 촉각 신호로부터 시각·청각적 특성을 복원할 수 있는 「인공감각(Artificial Sensation) 프레임워크」를 제안한다. 이는 기존의 감각 재현(reproduction) 중심 접근에서 벗어나, 감각 생성(generation)과 감각 상상(imagination)을 목표로 하는 새로운 미디어 패러다임이다. 즉, AI가 감각을 이해하고 상상할 수 있다면, 촉각이 없는 환경에서도 촉각을 느끼고, 소리 없는 공간에서도 촉각적 청각을 경험할 수 있다.

## II. 인공감각 프레임워크의 개념 및 구조

본 연구에서 제안하는 인공감각은 감각간의 양방향 상호생성(cross-modal generation)을 핵심으로 한다. 기존의 멀티미디어 시스템이 각 감각 채널을 병렬적으로 결합했다면, 인공감각은 감각 간 의미적 예측과 변환을 통해 하나의 감각이 다른 감각을 생성할 수 있는 자기완결적 감각 네트워크(self-sufficient sensory network) 구조를 목표로 한다.

## (1) 시청각→촉각: 센서리스 햅틱 생성

첫 번째 단계는 영상과 오디오만으로 햅틱 피드백을 생성하는 과정이

다. 인간은 물체의 충돌 장면이나 타격음, 질감 소리 등을 통해 접촉의 강도나 재질을 유추할 수 있다. 본 연구는 그림 1과 같이 이러한 인간의 인지적 추론을 모사하기 위해, 영상과 오디오 데이터를 입력으로 하여 촉각적 반응을 예측하는 멀티모달 인공지능 모델을 설계한다.

영상 스트림에서는 움직임의 크기, 속도, 가속도, 충돌 지점, 접촉 여부, 피사체 간 거리 등을 추출한다. 오디오 스트림에서는 음압, 주파수 스펙트럼, 지속 시간, 충격음의 파형, 소리의 종류를 분석한다. 이후 두 입력을 통합하여 [x, y, intensity, duration] 형태의 햅틱 벡터를 출력하는 신경망 구조(CNN+GRU, Temporal Convolution 등)를 적용할 수 있다. 이 모델은 촉각의 강도, 지속 시간, 공간적 위치 등 햅틱 피드백의 1차적 물리적 속성을 예측할 수 있으며, 그 결과물은 진동 모듈, 햅틱 수트, 또는 스마트폰 인터페이스를 통해 출력 가능하다.

이러한 구조는 기존의 센서 기반 햅틱 수집 과정을 완전히 제거하며, AI가 시청각 정보만으로 촉각적 상상력(haptic imagination)을 수행하는 형태로 작동한다. 즉, 인간이 보는 것과 듣는 것을 통해 느낄 수 있듯이, AI도 데이터를 통해 촉각을 예측할 수 있는 단계로 진입할 수 있다.

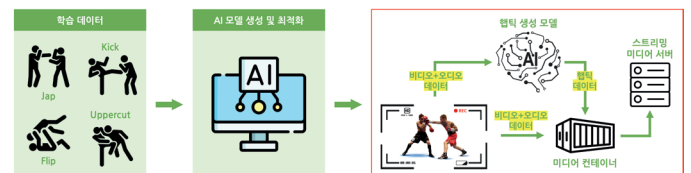


그림 1. 시청각→촉각: 센서리스 햅틱 생성 프레임워크 예시

## (2) 촉각 → 시청각: 햅틱 기반 감각 복원

본 연구에서 제안하는 촉각 기반 감각 복원은 햅틱 데이터의 시맨틱(semantic) 전송 구조를 핵심으로 한다. 기존의 접근이 진동 신호나 압력 분포와 같은 물리적 센싱 데이터 자체를 전송하는 방식이었다면, 본 프레임워크는 이러한 저차원 데이터를 직접 전송하지 않고, AI가 햅틱 입력을

통해 의미 단위(semantic token)를 추출하여 전송하는 방식을 채택한다. 예를 들어, 사용자가 물체의 표면을 긁거나 쓰다듬거나 두드리는 등 다양한 햅틱 행위를 수행하면,

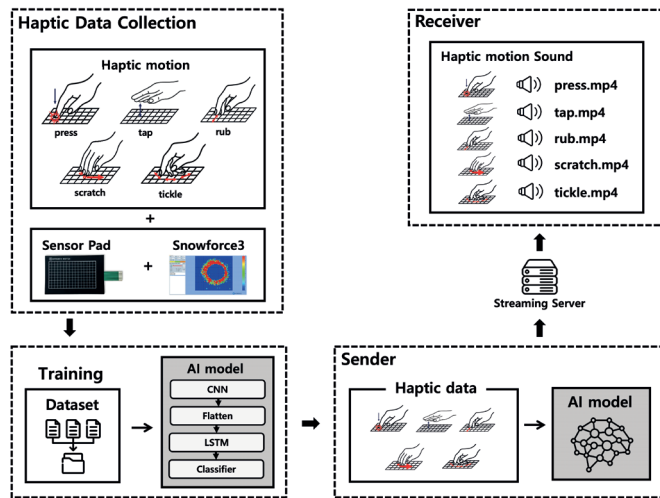


그림 2. 촉각 → 시청각: 햅틱 기반 감각 복원 프레임워크 예시

그림 2와 같이 송신 단의 인공지능 모델은 이를 'press', 'tap', 'rub', 'scratch', 'tickle' 등의 행위 단위로 분류하고, 그 위치와 맥락을 포함한 시맨틱 정보를 "사용자가 특정 위치를 쓰다듬고 있음" 또는 "금속 표면을 긁는 중"과 같은 의미 표현 형태로 압축하여 전송한다. 이때 전송되는 데이터는 원본 햅틱 파형이나 센서 스트림이 아니라, 행위의 의미를 표현하는 추상화된 감각 기호(sensory symbol)이다.

수신 단에서는 이러한 시맨틱 정보를 입력으로 받아, 해당 행위에 대응되는 시각적 장면(예: 손이 표면을 문지르는 동작)이나 청각적 효과(예: 긁힘 소리, 마찰음 등)를 생성 모델을 통해 복원한다. 즉, 수신자는 송신자가 느낀 감각을 동일한 물리 신호로 복제하는 대신, 그 의미적 맥락을 기반으로 새로운 감각 경험을 합성(synthesize)한다. 이는 감각을 데이터가 아니라 언어적 의미 단위로 다루는 새로운 감각 통신 구조이며, 실제 신호의 대역폭을 획기적으로 줄이면서도 감각적 몰입도는 유지할 수 있다는 장점을 가진다.

결과적으로, 제안된 접근은 햅틱 데이터를 물리적 신호의 복제에서 감각의 언어적 재구성으로 전환시키며, 감각의 본질을 의미적 행위 표현으로 추상화하는 인공감각(AI-driven semantic perception) 프레임워크의 핵심을 이룬다.

### III. 기대효과 및 응용 가능성

제안하는 인공감각 프레임워크는 감각을 물리적 신호가 아닌 의미 단위로 전송함으로써, 실감형 미디어 전송의 새로운 가능성을 제시한다. 기술적으로는 센서 없이도 AI가 시청각으로부터 촉각을 생성하고, 반대로 촉각 행위를 의미적으로 해석해 시청각 경험으로 복원함으로써 데이터 전송량을 줄이면서도 높은 몰입감을 유지할 수 있다.

이 접근은 실감형 방송, OTT, 원격 협업, 의료재활, 교육용 XR 등 다양한 산업 분야에 적용 가능하며, 기존 콘텐츠에도 후처리 형태로 쉽게 확장될 수 있다. 또한 감각을 데이터가 아닌 언어적 정보로 다루는 감각 통신(sensory communication) 개념을 실현함으로써, 차세대 초실감 미디어와 인공지능 감각지능 연구의 기반 기술로 발전할 잠재력을 지닌다.

### IV. 결론

본 연구는 인공지능을 기반으로 감각을 상호 추론하고 생성할 수 있는 인공감각(Artificial Sensation) 프레임워크를 제안하였다. 이 프레임워크는 시청각과 촉각 간의 상호생성(cross-modal generation)을 통해, 하나의 감각이 다른 감각을 예측하고 의미적으로 변환함으로써 감각의 통신(sensory communication) 개념을 구체화한다.

특히, 제안된 구조는 송신자가 물리적 햅틱 데이터를 직접 전송하는 대신, "무엇을 어떻게 느끼고 있는가"라는 시맨틱 정보를 인코딩하여 전송하고, 수신자는 이를 해석하여 대응되는 시각적 또는 청각적 경험을 복원하는 방식을 취한다. 이러한 접근은 감각을 물리적 신호로 복제하는 기존 방식에서 벗어나, 감각의 의미를 생성하고 재구성하는 새로운 패러다임을 제시한다.

결과적으로 인공감각은 감각을 데이터로 다루는 기존 멀티미디어 구조를 넘어, AI가 감각을 '이해하고 공유할 수 있는' 존재로 진화하는 가능성을 보여준다. 이는 단순한 기술적 확장을 넘어, 인간의 지각 메커니즘을 디지털적으로 해석하고, 감각 간 경계를 초월한 상호지각적(Inter-perceptual) 미디어 경험을 가능하게 할 것으로 기대한다.

### ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [25ZC1100, 초실감 입체공간 미디어·콘텐츠 원천기술 연구]

### 참 고 문 헌

- [1] F. Danieau, A. Lecuyer, P. Guillotel, J. Fleureau, N. Mollet and M. Christie, "Enhancing Audiovisual Experience with Haptic Feedback: A Survey on HAV," *IEEE Transactions on Haptics*, vol. 6, no. 2, pp. 193-205, April-June 2013.
- [2] 전승엽, 이용준, 정승혁, 김성훈, 유동호, "ATSC 3.0을 위한 햅틱: ATSC A/380 권고안 소개 및 도전과제," *방송공학회논문지*, 제29권, 제6호, pp. 954-963, 2024.
- [3] Y. Lee and D. You, "Demonstration of WebRTC-based Video Conferencing with Tactile Sensations," *2024 IEEE Consumer Life Tech (ICLT)*, Sydney, Australia, 2024.