

# 한국음악정보학회 2026

## AI 피아노 연주 모델은 ‘표현’을 어떻게 학습하는가?

### - RENCON 2025 참가 모델의 구조 분석과 연주학적 고찰 -

익명 저자 (Anonymous Authors)

익명 소속 (Anonymous Affiliation)

anonymous@ksmi.kr

#### 1. 초록

AI가 과연 프로 연주자처럼 표현력 있는 연주를 할 수 있을까? 본 연구는 2025년에 열린 표현적 연주 렌더링 경연대회 RenCon에 참가한 AI 피아노 연주 모델 7개를 대상으로, 각 모델이 인간 연주의 핵심 표현 요소(템포, 다이내믹, 아티큘레이션, 페달링 등)를 어떤 방식으로 데이터화하고 알고리즘화하는지 구조적으로 비교하고, 분석한다. 딥러닝 기반 모델부터 규칙 기반 모델까지 다양한 접근법을 연주학적 시각에서 검토함으로써, AI가 ‘표현’을 학습 가능한 변수로 전환하는 과정의 의미와 한계를 고찰한다. 나아가 표현의 주체가 누구인지라는 미학적 질문을 제기하며, 향후 청취 평가 및 포스트휴머니즘 이론과의 접목을 통한 후속 연구 방향을 제시한다.

#### 2. 연구의 출발점

AI는 작곡, 편곡, 채보를 넘어 이제 ‘연주’의 영역에까지 진입하고 있다. 야마하의 연주 표현 렌더링 연구 중 글렌 골드 스타일을 재현한 *Dear Glenn* 프로젝트는 단순한 기술 실험을 넘어, AI 연주가 실제 공연 맥락에서 활용될 수 있음을 보여준다 [1]. 이러한 흐름 속에서 ‘표현적 연주 모델링’(Expressive Performance Modeling, EPM) 연구는 빠르게 성장해왔다 [2].

이 맥락에서 주목할 만한 사실이 있다. 2002년 RenCon의 창시자 히라가(Hiraga)는 “반세기 안에 연주 렌더링 시스템이 쇼팽 콩쿠르에서 우승할 것”을 목표로 제시했다 [3]. 그로부터 23년이 지난 2025년, 실제로 개최된 RenCon 2025에서 인간 연주는 여전히 4.4점으로 최고 AI 모델(3.62점)을 크게 앞질렀다. 이 간극은 단순한 기술적 미완성이 아니라, ‘연주’라는 행위의 본질이 얼마나 복합적인지를 드러낸다.

그러나 기존 EPM 연구의 대부분은 모델의 성능 향상이나 기술적 비교에 집중되어 있으며, 연주의 본질(해석, 표현, 신체성)에 대한 음악학적 논의는 상대적으로 부족하다. 본 연구는 이 공백을 메우고자, 동일한 과제와 평

가 체계 아래 다양한 접근법의 모델들이 경연한 RenCon 2025를 분석 대상으로 삼는다.

핵심 연구 질문은 다음과 같다. (1) AI 피아노 연주 모델들은 인간 연주의 표현적 요소를 어떤 구조적 방식으로 구현하는가? (2) 표현을 수량화하고 학습하는 과정에서 무엇이 포착되고, 무엇이 소실되는가?

#### 2. 연구 대상: RENCON 2025

RenCon(International Performance Rendering Contest)은 악보로부터 표현적 연주를 생성하는 자동화 시스템을 겨루는 대회로, 2002년부터 2011년까지 총 6회 개최된 후 14년 만에 ISMIR 2025와 함께 재개되었다 [3]. 대회는 시작 당시부터 연주 렌더링 시스템의 평가 방법을 확립하는 것을 목표로 했으며, 시대에 따라 튜링 테스트 도입(2004), 실시간 렌더링 방식 채택(2008), 이중 평가 체계 구축(2011) 등 평가 방식을 지속적으로 발전시켜 왔다.

2025년 대회에서는 본선에 8개 모델이 참가하였으며, 비교 지표로 인간 연주도 함께 평가되었다. 인간 연주는 4.4점(5점 만점)으로 전체 1위를 차지하였고, 모델 중에서는 VirtuosoNet이 3.62점으로 최고점을 기록하였다. 본 연구는 기술보고서가 제출된 7개 모델을 분석 대상으로 삼는다: VirtuosoNet, Contin-U, YQX+, Automated Expressive MIDI Renderer(AEMR), Director Musices(DM), midihum, PerTok Transformer.

#### 3. 연구내용: 모델 구조 비교 분석

##### 3.1 입출력 방식 및 렌더링 파이프라인

7개 모델의 입출력 형태와 렌더링 방식을 표로 정리하였다.(표 1).

##### 3.2 표현 파라미터 구현 방식

- 계층적 구조 모델링: VirtuosoNet은 마디(measure) 수준과 음표(note) 수준의 이중 계층 구조를 통해 장기적 음악 구조를 반영한다. RNN 기반의 Hierarchical Attention Network와 Conditional Variational Autoencoder

모델	입력	출력	방식
VirtuosoNet	MusicXML	MIDI	RNN + CVAE
Contin-U	이미지	오디오	cross modal transformer
YQX+	MusicXML	MIDI	확률적 생성 모델
AEMR	MusicXML	MIDI	transformer
DM	MusicXML/ MIDI	MIDI	rule-based
midihum	MIDI	MIDI	XGBoost
PerTok	MusicXML	MIDI	비자기회귀 transformer

표 1. RenCon 2025 참가 모델 유형 비교

를 결합하여, 동일한 악보에서 다양한 연주 해석이 가능하다는 점을 latent space에서 모델링한다 [4].

• **확률적 표현 생성:** YQX+는 단일 결정론적 출력 대신 확률분포 기반의 다양한 표현 가능성을 탐색한다. GMM,  $\beta$ -VAE, Flow-Matching 등을 통합한 프레임워크로, 동일 악보에서 스타일이 다른 여러 연주를 생성할 수 있다.

• **종단형(end-to-end) 방식:** Contin-U는 악보 이미지를 RVQ 토큰으로 변환하여 오디오를 직접 생성하는 cross-modal transformer다. symbolic 표현 파라미터를 별도로 예측하지 않고, 모델 내부에서 암묵적(implicit) 표현 모델링을 수행한다.

• **규칙 기반 해석 가능 모델:** Director Musices는 프레이즈 아치 규칙, 다이내믹 SVR 모델, 최종 리타르단도 모델 등 독립적으로 조절 가능한 규칙 집합으로 구성된다. 답러닝 모델과 달리 각 표현 결정의 근거가 투명하게 드러나며, RenCon 2002년부터 꾸준히 참가해온 유일한 모델이기도 하다.

### 3.3 연주학적 시각에서의 분석

연주학(performance studies)의 관점에서 볼 때, 각 모델의 구조적 선택은 ‘연주란 무엇인가’에 대한 서로 다른 전제를 내포한다. VirtuosoNet의 계층적 구조는 음악 형식과 프레이징에 대한 분석적 이해를 전제하며, YQX+의 확률적 모델링은 연주 해석의 다양성과 불확정성을 인정한다. 반면 Contin-U의 종단형 접근은 표현 파라미터를 명시적으로 정의하지 않음으로써, ‘표현’이 분해 가능한 요소들의 합인지에 대한 근본적 질문을 제기한다.

규칙 기반 DM은 Friberg 등의 KTH 연주 규칙에 기반하며 [5], 이는 연주 표현이 원칙적으로 설명 가능하다는 인지과학적 관점을 반영한다. 이처럼 서로 다른 모델들은 각각 다른 연주 이론과 암묵적으로 대화하고 있다.

## 4. 논의 및 향후 연구

### 4.1 무엇이 포착되고 무엇이 소실되는가

현재 AI 연주 모델들은 템포 곡선, velocity 분포, 온셋 편차 등 수량화 가능한 표현 요소는 상당 수준으로 구현하는 반면, 연주자의 신체적 제스처에서 비롯되는 음색 변화, 즉흥적 해석 결정, 청중과의 실시간 상호작용 등은 현재의 프레임워크 안에서 포착되기 어렵다. 이는 단순한 기술적 한계가 아니라, ‘표현’을 학습 가능한 변수로 환원하는 과정 자체에 내재한 구조적 간극이다.

더 나아가, 표현의 주체가 누구인가라는 질문이 제기된다. AI 모델이 생성한 연주에서 표현의 책임은 모델 자체에 있는가, 학습 데이터를 제공한 원래 연주자에게 있는가, 아니면 모델을 설계한 연구자에게 있는가? 이는 포스트휴머니즘 이론(특히 Hayles의 탈신체화(disembodiment) 논의나 Braidotti의 포스트휴먼 주체성 개념)과 접목될 때 더욱 풍부하게 탐구될 수 있다 [6, 7].

### 4.2 향후 연구 방향

본 연구는 석사 학위논문의 일환으로 진행 중이며, 다음과 같은 후속 연구를 계획하고 있다.

우선 RenCon 2025의 청취 평가 데이터와 필자의 분석을 토대로, 인간이 어떤 표현 요소에서 특정 모델의 연주를 자연스럽다고 인식하는지 정성적, 정량적으로 분석한다. 이를 통해 ‘사람답다’는 판단의 기준이 무엇인지를 규명할 것이다. 또한 Rink [8]의 연주 이론 등 기존 연주학 논의를 AI 연주 모델 분석에 적용하여, 기술과 음악학 사이의 새로운 대화를 시도할 것이다. 더 나아가 AI가 연주라는 신체적·해석적 행위를 시도한다는 것이 음악 존재론적으로 무엇을 의미하는지를, 포스트휴머니즘 이론 틀 안에서 심층적으로 고찰한다. 궁극적으로 본 연구는 AI 연주 모델이 촉발하는 ‘연주의 재정의’ 문제를 음악학의 중심 의제로 끌어올리는 데 기여하고자 한다.

## 5. 맺음말

앞으로 본 연구는 RenCon 2025에 참가한 7개 AI 피아노 연주 모델의 구조를 비교 분석함으로써, AI가 ‘표현’을 어떻게 데이터화·알고리즘화하는지를 연주학적 관점에서 검토할 것이다. 각 모델이 채택한 구조적 접근법은 연주 표현에 대한 서로 다른 이론적 전제를 내포하며, 이는 단순한 기술 비교를 넘어 ‘표현이란 무엇인가’라는 근본 질문과 닿아 있다. 2002년 RenCon이 품었던 야망(쇼팽 콩쿠르를 제패하는 AI)은 아직 실현되지 않았지만, 그 간극을 음악학적으로 해명하는 작업이야말로 다음 단계의 핵심 과제다. AI 연주 연구는 기술적 성과 지표만이 아니라 음악학·미학·철학과의 교차적 논의를 통해 더욱 깊어질 수 있으며, 본 연구는 그 출발점을 마련하고자 한다.

## 6. 참고문헌

- 143 [1] Yamaha Corporation, “Glenn Gould AI performs for  
144 the first time ever at the Ars Electronica festival,” 2019,  
145 [https://www.yamaha.com/en/stories/new-values/  
146 dear-glenn/](https://www.yamaha.com/en/stories/new-values/dear-glenn/).
- 147 [2] C. Hernandez-Olivan, J. Hernandez-Olivan, and J. R.  
148 Beltran, “A survey on artificial intelligence for music  
149 generation: Agents, domains and perspectives,” *arXiv  
150 preprint*, 2022, arXiv:2210.13944.
- 151 [3] R. Hiraga, K. Hirata, and H. Katayose, “The second  
152 rencon: Performance contest, panel discussion, and the  
153 future,” in *Forum on Information Technology*, 2002.
- 154 [4] D. Jeong, T. Kwon, Y. Kim, and J. Nam, “Graph neural  
155 network for music score data and modeling expressive  
156 piano performance,” in *Proc. of the 36th Int. Conf. on  
157 Machine Learning (ICML 2019)*, 2019, pp. 3060–3070.
- 158 [5] A. Friberg, R. Bresin, and J. Sundberg, “Overview of  
159 the KTH rule system for musical performance,” *Ad-  
160 vances in Cognitive Psychology*, vol. 2, no. 2–3, pp.  
161 145–161, 2006.
- 162 [6] N. K. Hayles, *How We Became Posthuman: Virtual  
163 Bodies in Cybernetics, Literature, and Informatics*.  
164 Chicago: University of Chicago Press, 1999.
- 165 [7] R. Braidotti, *The Posthuman*. Cambridge: Polity  
166 Press, 2013.
- 167 [8] J. Rink, Ed., *Musical Performance: A Guide to Under-  
168 standing*. Cambridge: Cambridge University Press,  
169 2002.

## 7. AI 사용 명세서

- 171 기술적 내용의 이해와 LaTeX 문법 학습에 Claude (An-  
172 thropic의 Sonnet 4.6 모델)를 사용하였다.