

A Historical Review of AI-Based Music Composition: From Local Pattern Prediction to Holistic Audio Generation

Byunghyun Ban^{1,2}¹Gyeongbuk National University, Department of Music, ²NaNa Lab

ABSTRACT

AI-based music composition has evolved from explicit rule encoding and local symbolic prediction to end-to-end generation of integrated musical audio. This historical review examines representative systems that either introduced a distinct generative paradigm or achieved practical deployment. It traces the progression from rule-based and stochastic models through recurrent networks and Transformers to waveform, latent-audio, diffusion, flow-matching, hybrid, and multimodal approaches. Early models reproduced local transitions among notes, chords, and rhythms but struggled with long-range form and thematic development. Recurrent and attention-based architectures expanded temporal context, yet largely remained within symbolic prediction. Audio-level models shifted the generative target from musical instructions to sound itself, enabling joint modeling of structure, performance, timbre, texture, and acoustic realization. The review argues that the key historical transition was not simply an increase in model capacity, but a change in representational scope: from decomposing music into human-defined units to learning high-dimensional relationships more holistically. Current systems still face unresolved problems in long-term planning, fine-grained control, editability, and human-AI collaboration. Nevertheless, their development reflects a broader movement from automating compositional procedures toward modeling increasingly integrated stages of musical ideation and sound formation.

• Keyword : compositional AI, AI-based composition, generative AI, pattern reconstruction

AI기반 작곡 기술의 역사적 고찰: 국소 패턴 예측에서 총체적 오디오 합성으로

반병현^{1,2}¹국립경북대학교 공연예술음악과, ²나나랩 (NaNa Lab)

요 약

AI 기반 음악 작곡 기술은 명시적 규칙의 부호화와 국소적 심볼 예측에서 통합된 음악 오디오를 생성하는 end-to-end 방식으로 발전해 왔다. 본 연구는 새로운 생성 패러다임을 제시했거나 실제 서비스로 구현된 대표 기술을 중심으로 이러한 발전 과정을 역사적으로 고찰한다. 규칙 기반 및 확률 모델에서 순환신경망과 트랜스포머를 거쳐, 파형·잠재 오디오·확산·플로 매칭·하이브리드 및 멀티모달 모델로 이어지는 기술적 전환을 분석하였다. 초기 모델은 음표, 화음, 리듬 사이의 국소적 전이 패턴을 재현할 수 있었으나 장기적인 형식과 주제 전개를 구성하는 데 한계를 보였다. 순환 및 어텐션 기반 모델은 시간적 문맥을 확장했지만 주로 심볼릭 예측의 범위에 머물렀다. 반면 오디오 수준의 모델은 생성 대상을 음악적 지시에서 소리 자체로 전환함으로써 구조, 연주, 음색, 질감 및 음향적 구현을 통합적으로 학습하기 시작하였다. 이러한 발전의 핵심은 단순한 모델 용량의 증가가 아니라, 인간이 정의한 음악 단위를 분해하여 처리하는 방식에서 고차원적 관계를 총체적으로 학습하는 방식으로 표현 범위가 변화한 데 있다. 현재의 시스템은 장기적 구조 설계, 세밀한 제어, 편집 가능성 및 인간과 AI의 협업 측면에서 여전히 한계를 지니지만, 그 발전 과정은 작곡 절차의 자동화에서 음악적 아이디어와 소리 형성의 통합적 모델링으로 이동하고 있음을 보여준다.

• 중심어 : 작곡 인공지능, 음악 인공지능, 생성형 인공지능, 패턴 인식

¹Corresponding author: bhban@nanalab.kr

Received 2026.05.02; Revised 2026.06.19; Accepted 2026.07.13

www.kci.go.kr



© 2026 by Korea Association of Artificial Intelligence Education. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.