

AI 커버곡의 기술적 구조와 음성 변환 기술의 확장

한국저작권위원회
정보기술팀
문수연
2025. 12. 15.

보고서 요약

AI 커버(Cover)곡은 인공지능 기술을 활용하여 기존 음악의 보컬을 특정 가수의 목소리로 변환하거나 원곡의 스타일을 새롭게 재해석한 음악을 의미한다. 제작 과정에는 딥러닝 기반 음성 모델링, 복제, 변환, 노래 보컬 생성에 활용되는 음성 합성 등 다양한 기술이 결합된다.

핵심 기술인 음성 변환은 원본 노래의 멜로디와 리듬을 유지한 채 음색만 특정 인물의 목소리처럼 바꿔주는 기능을 수행하며, AI 커버곡의 현실감과 완성도를 결정하는 주요 요소로 작용한다. 이 기술은 음악 제작뿐 아니라 더빙, 가상 캐릭터, 장애인 보조 기술 등 여러 산업으로 확장되고 있다.

그러나 AI 커버곡은 저작권·초상권 등 권리 구조가 아직 명확히 정립되지 않아 주요 음원 플랫폼에서는 공식적으로 제한되는 경우가 많다. 기술적·윤리적 위험에 대한 우려도 존재하면서, 관련 제도 정비와 기술 활용 기준 마련이 요구된다. 제도적 기반이 마련될 경우, 앞으로 산업과 창작 영역 모두에서 긍정적 발전 가능성을 가질 것으로 기대된다.

1. AI 커버곡의 개념과 시장 동향

1) AI 커버곡의 개념

최근 가수 목소리를 인공지능 기술로 흉내 낸 노래 영상이 유튜브를 통해 확산되고 있다.

AI 커버(Cover)곡은 인공지능 기술을 활용하여 기존 음악의 보컬을 특정 가수의 목소리로 변환하거나 원곡의 스타일을 새롭게 재해석한 음악을 의미한다. 이는 기술적으로 기존 곡을 기반으로 하되 전혀 다른 음색과 감성으로 재구성된 콘텐츠이다.¹⁾

AI 커버곡이 널리 활용되면서, 그 제작 방식과 기존 커버곡과의 차이를 이해하려는 관심도 함께 높아지고 있다. 이어지는 내용에서는 AI 커버곡이 어떤 기술로 만들어지는지, 어떤 점에서 기존 방식과 구별되는지, 그리고 생성형 AI가 음악 분야에서 어떤 흐름을 만들어 내고 있는지를 간단히 살펴본다.

2) 기존 커버곡과 AI 커버곡의 차이

일반적인 커버곡이 다른 가수의 곡을 자신만의 음악 스타일로 재해석해 직접 부르는 방식이라면, AI 커버곡은 A가 부른 노래를 마치 B가 부른 것처럼 만드는 것이다.²⁾ 두 방식은 모두 원곡의 멜로디와 가사를 기반으로 한다는 공통점이 있으나, 제작 과정과 사용되는 기술, 보컬 음색의 재현 방식에서 뚜렷한 차이가 존재한다. 특히 AI 커버곡은 음성 변환과 합성 기술을 기반으로 특정 음색을 정밀하게 모사할 수 있다는 점에서 기존 커버곡과는 다른 기술적 특성을 가진다. 아래 표는 이러한 차이를 주요 요소별로 비교한 것이다.

| 기존 커버곡 vs AI 커버곡: 요소별 비교

구분	실연 주체	음색 재현 방식	생산 비용·속도
기존 커버곡	사람이 직접 노래를 부르는 실연	가수가 최대한 비슷하게 모창하더라도, 생리적 한계로 인해 동일한 음색 구현은 불가능	녹음실, 보컬, 장비 등이 필요하고 반복 녹음이 수반
AI 커버곡	딥러닝 기반 음성 합성 모델이 보컬을 생성	학습 데이터를 통해 특정 화자의 음색·발성 특징을 수치화하고, 동일한 음색을 재현	텍스트, 원곡 음원만으로도 다양한 버전의 커버를 빠르게 생성 가능

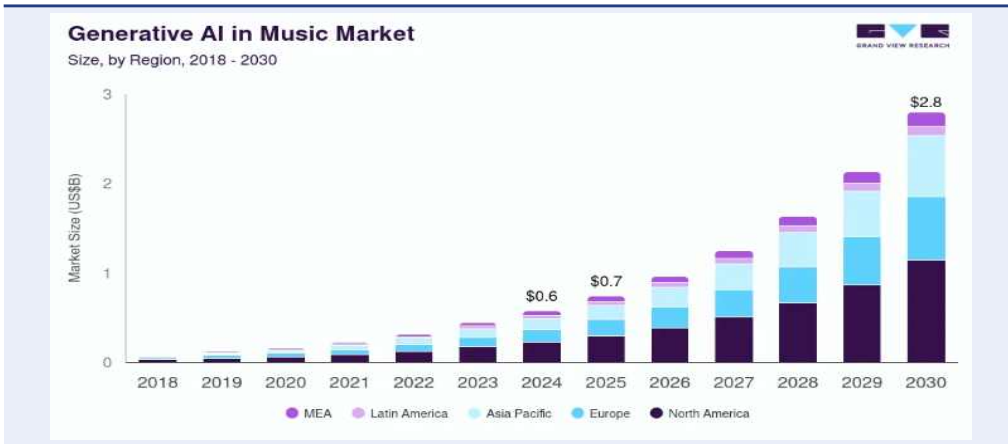
1) KCI, "AI 커버곡의 저작권법적 쟁점과 실무적 접근 방안", 2025.

2) 한국콘텐츠진흥원, "2024 음악산업백서", 2024.09.30.

3) 음악 분야에서의 생성형 인공지능 확산과 시장 동향

AI 커버곡은 생성형 인공지능 기반 음악 기술의 활용 사례 중 하나로, 이러한 기술이 포함된 생성형 인공지능 음악 시장은 최근 빠르게 성장하고 있다. 음악 분야의 글로벌 생성 AI 시장 규모는 2023년 4억 4천만 달러(원화 약 6,477억 2,400만 원)³⁾ 추정되었으며, 2030년 예상 시장 규모는 27억 9,470만 달러(원화 약 4조 1,140억 7,800만 원)로 예상된다. 이는 2023년~2030년 연평균 성장률(CAGR)이 30.4%에 해당하며 **생성형 AI 음악 기술이 가파른 성장세를 가진 신흥 시장임을 보여준다.**

| 글로벌 생성형 인공지능 음악 시장 규모 및 전망



※ 출처: Grand view research, "Generative AI in Music Market (2024 - 2030)"

3) 2025.12.08. 기준의 하나은행 환율 최초 매매 기준율을 적용. 해당 목차 이하 동일.

2. AI 커버곡을 구성하는 기술적 기반

AI 커버곡 기술은 인공지능(AI)을 활용해 특정 인물의 목소리 특징을 학습하고 이를 다른 노래에 적용하여 새로운 보컬을 생성하는 기술이다. 딥러닝 기반 알고리즘을 통해 음색·발음·표현 방식 등 목소리의 세부 특성을 분석·모델링함으로써, 원곡과는 다른 목소리로 노래가 재해석된 결과물을 만들어낼 수 있다.

1) AI 커버곡 제작 원리

연예인이 부른 음원, 방송에서 한 대사와 말이 담긴 음원 파일 등을 AI 딥러닝에 입력하면 AI는 연예인의 목소리 데이터값으로 딥러닝을 통해 목소리의 특성을 추출한다. 이후, 음원을 넣고 AI에 변환 명령을 내리면 AI가 음원 특성 데이터를 연예인 목소리 특성 데이터로 전환하여 해당 연예인의 AI 커버곡 제작이 완성된다.

특히 AI가 딥러닝 과정에서 숨소리까지 포착해 학습하기 때문에 모르고 들으면 당사자가 부른 것과 같은 착각을 일으킬 정도로 정교하며, 데이터를 노래에 덮어 씌우는 원리로 음치도 AI를 활용하면 가수처럼 노래를 부를 수 있는 것이 특징이다.⁴⁾

| AI 커버곡 제작 원리



※ 출처: 한국과학기술기획평가원, "AI 커버곡, 생성 AI 기술의 새로운 시장을 형성"

2) AI 커버곡 핵심 기술

AI 커버곡 제작 과정은 크게 세 단계로 구분된다. 첫째, **목소리 학습 단계**에서는 딥러닝 알고리즘을 사용하여 특정 가수의 목소리 데이터를 학습하고, 음색·음역·표현방식·발음 등 다양한 특징을 분석한다. 둘째, **목소리 합성 단계**에서는 학습된 모델을 기반으로 새로운 목소리 트랙을 생성하고 원본 노래의 멜로디, 리듬, 가사에 맞춰 목소리를 합성한다. 마지막으로 **음악 제작 단계**에서는 이렇게 합성된 목소리 트랙을 기반으로 하나의 음악 작품으로 완성한다.⁵⁾

4) 한국과학기술기획평가원, "AI 커버곡, 생성 AI 기술의 새로운 시장을 형성", 2024.03.12.

5) 한국저작권보호원, "AI 커버곡과 저작권 침해", 2024.07.23.

앞서 살펴본 제작 원리를 바탕으로, AI 커버곡 제작에는 여러 핵심 기술이 결합되어 활용된다. 딥러닝 기반 음성 모델링(Speaker Modeling)과 음성 복제(Voice Cloning), 보컬의 음색을 변환하는 음성 변환(Voice Conversion, VC) 기술이 핵심적으로 사용된다. 여기에 노래 보컬 생성을 위한 노래 음성 합성(Singing Voice Synthesis, SVS)과 후처리 기술이 함께 적용되어 최종 결과물을 완성한다. 아래 표는 AI 커버곡 제작 과정별 핵심 기술과 그 역할을 정리한 것이다.

| AI 커버곡 제작 과정별 핵심 기술 정리

구분(단계)	주요 기술	기술 내용
목소리 학습 단계	딥러닝 기반 음성 모델링 (Speaker Modeling)	특정 인물의 음색·발음·억양 등 목소리 특징을 분석하여 AI가 사용할 수 있는 음색 표현으로 모델링하는 기술
	음성 복제 (Voice Cloning)	학습된 대상 목소리 모델을 기반으로 실제 특정 인물의 목소리처럼 들리는 합성 음성을 생성하는 기술
목소리 합성 단계	음성 변환 (Voice Conversion, VC)	원본 노래의 멜로디·리듬을 유지하면서 보컬의 음색만 특정 인물의 목소리처럼 변환하는 기술
	노래 음성 합성 (Singing Voice Synthesis, SVS)	가사·멜로디 정보를 바탕으로 노래 보컬을 합성하되, AI 커버곡에서는 음성 변환 과정에서 추출된 멜로디·리듬 정보를 재구성하는 데 보조적으로 사용되는 기술
음악 제작 단계	음성 특징 분석 및 재구성 (Feature Extraction & Reconstruction)	피치·리듬·발성 표현 등 노래 특유의 음성 특징을 분석하고 자연스럽게 재구성하여 품질 높은 보컬을 만드는 기술
	AI 기반 제작 도구 활용 (RVC 등 오픈소스·상용 툴)	생성된 보컬을 반주와 결합하고 편집·믹싱·후처리를 수행하여 최종 음악으로 완성하는 도구

3. 음성 변환 기술의 확장과 활용: AI 커버곡을 넘어

앞서 살펴본 AI 커버곡 제작의 핵심 기반 기술인 음성 변환(Voice Conversion, VC) 기술은 음악 콘텐츠를 넘어 다양한 산업으로 활용 범위를 넓혀가고 있으며, 생성형 인공지능의 발전과 함께 성능과 적용 가능성도 빠르게 향상되고 있다.

1) 음성 변환 기술⁶⁾

음성 변환 기술은 특정 개인의 목소리를 다른 사람의 목소리로 변환하는 기술이다. 현재 다양한 분야에서 활용되고 있으며, 주요 적용 분야로는 음성 합성 및 텍스트 변환, 음성 감정 변화, 음성 변조 및 보안, 음성 관련 연구 등이 있다.

아래 표는 음성 변환 기술의 기본 원리와 대표적인 변환 기술 및 모델을 정리한 것이다.

I 음성 변환 기술 개요(원리·기술·모델)

구분		내용
기본 원리	특징 벡터 추출	입력 음성과 목표 음성의 특징 벡터를 추출 음성의 주파수, 음의 높낮이, 강약 등을 표현하는 정보를 담고 있음
	통계적 모델링	추출한 특징 벡터를 통계적 모델에 적용하여 음성 변화를 수행 입력 음성의 특징 벡터를 목표 음성의 특징 벡터로 매핑하는 방법을 학습
	목표 음성 생성	통계적 모델을 사용하여 입력 음성의 특징 벡터를 목표 음성의 특징 벡터로 변환하여 생성된 목표 음성은 입력 음성과 유사한 발음과 억양을 가지면서도 목표 음성의 특징을 반영
주요 변환 기술 및 모델	스타일 제어	음성 변환은 목표 음성의 스타일을 다양하게 제어 가능 이를 통해 다양한 발화 스타일을 생성
	StarGAN-VC	GAN(Generative Adversarial Networks) 아키텍처를 ⁷⁾ 사용하여 실시간 음성 합성을 가능하게 하는 모델로 여러 목소리로의 변환을 비병렬 방식으로 수행
	AutoVC	AutoEncoder의 BottleNeck 구조를 ⁸⁾ 사용하여 발화자와 내용 정보를 분리하는 방식으로, 이를 통해 훈련 데이터에 없는 발화자의 목소리 변환도 고품질로 생성

6) 한국저작권보호원, "AI 커버곡과 저작권 침해", 2024.07.23.

7) GAN 아키텍처(생성적 적대 신경망, Generative Adversarial Networks): 생성자와 판별자 두 개의 신경망이 서로 경쟁하며 학습하는 인공지능 구조

8) 네트워크에서 정보량이 가장 제한된 중간 계층으로, 입력 데이터의 핵심 특성을 압축·추출하도록 설계된 구조

2) 음성 변환 기술의 주요 활용 분야⁹⁾

음성 변환 기술은 성별·나이·언어·음색 등 다양하게 목소리 변환이 가능하여 콘텐츠 제작·접근성 강화·개인화 서비스 등 여러 분야에서 활용되고 있다. 이러한 기술적 특성은 엔터테인먼트뿐 아니라 고객 응대·더빙·안내 등으로 적용 범위를 넓히고 있다. 아래 표는 음성 변환 기술이 실제로 활용되고 있는 주요 분야를 요약한 것이다.

I 음성 변환 기술 활용 영역 요약

구분		기술 내용
미디어	음악	가수의 목소리를 변환해 다양한 음색을 실험하거나 인공지능 가수의 목소리를 생성하는 데 활용
	더빙 및 성우	영화나 애니메이션에서 특정 배우의 목소리로 다른 언어로 더빙하거나 특정 캐릭터의 음성을 재현하는 데 사용
	시각 장애인	음성으로 변환해 주는 스크린 리더 소프트웨어를 사용하여 웹사이트, 문서, 이메일 등을 읽을 수 있게 도와줌
장애인 지원	청각·발음 장애인	음성 텍스트 변환 기술을 사용해 대화 등의 내용을 실시간으로 텍스트로 변환해 의사소통을 할 수 있도록 도와줌
	신체 장애인	보완 대체 보조 기구(터치스크린, 스위치 등) 활용 및 음성명령을 통해 전자기구나 스마트홈을 제어하여 독립적 생활이 가능
	의료 분야	의료 기록을 음성으로 입력하거나, 환자와의 대화를 기록 및 분석
기타	정보 보호	음성을 변조하여 실제 목소리를 보호하고, 개인정보 유출 위험 감소
	가상 비서	사용자가 선호하는 목소리로 설정하여 친근하고 개인화된 경험 제공
	교육 분야	언어 학습 시 발음을 교정하거나, 음성 합성 기술을 이용해 교재를 오디오 형태로 제공함으로써 다양한 학습 방식을 지원

위와 같이 다양한 활용에도 불구하고, 아직 해결해야 할 문제점도 존재한다. 현재 음성 변환 기술은 목소리를 변화하는 과정에서 자연스러움을 유지하는 데 어려움을 겪고 있으며, 특히 감정 표현이나 섬세한 음색 변화 과정에서 인공적인 느낌을 줄 수 있다. 또한 딥페이크 기술과 결합하여 가짜 뉴스 제작·보이스 피싱 등 범죄에 악용될 가능성이 제기되며, 개인정보 보호, 저작권 침해, 인간 존엄성 훼손 등 윤리적 우려도 함께 증가하고 있다.

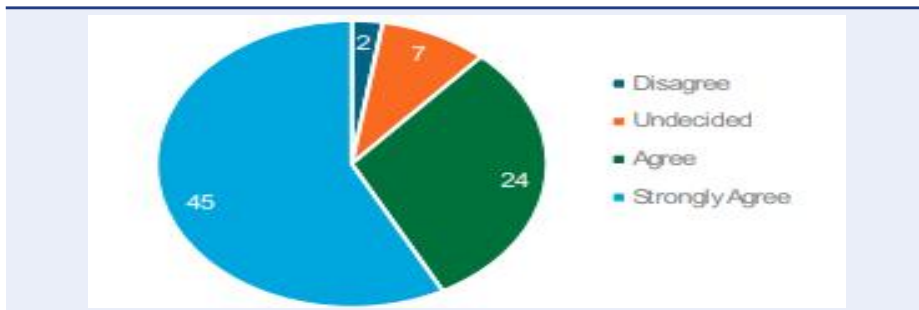
9) 한국저작권보호원, "AI 커버곡과 저작권 침해", 2024.07.23.

4. 시사점

AI 커버곡은 주로 유튜브, 틱톡 같은 개방형 플랫폼에서 다수 업로드되고 쉽게 접근 가능한 반면 멜론, 스포티파이 등 주요 음악 스트리밍 플랫폼에서는 저작권 및 기술적 제약으로 인해 공식적인 등록이 제한되고 있다. 이는 AI 커버곡과 관련된 법적 쟁점이 아직 정리되지 않았음을 보여주며, 권리자 보호와 기술 발전 사이의 균형을 어떻게 조정할 것인지에 대한 논의가 필요한 상황임을 시사한다.¹⁰⁾

AI 커버곡에 대한 법적 보호 필요성 인식을 확인하기 위해 실시한 조사 결과에서도 응답자의 과반수 이상인 45명(57.7%)이 법적 보호 조치가 필요하다고 답하였다. 이는 현재의 법적·제도적 장치가 AI 기반 음악 생성 기술이 야기하는 저작권, 윤리, 예술성 침해 문제 등을 규율하고 통제하기에 충분하지 않다는 사회적 인식을 제시하며, 정책적 대응의 필요성을 뒷받침하는 실증적 근거가 된다.

| AI 커버곡에 대한 법적 보호조치 필요성 인식



※ 출처: KCI, "AI 커버곡의 공정 이용에 관한 연구", 2024.12.

음성 변환 기술의 발전은 창의적 콘텐츠 제작에 새로운 가능성을 열어주지만, 이를 안전하게 활용하기 위한 제도적·기술적 장치 마련이 필요하다. 앞으로 권리 보호 기준과 기술 활용 원칙이 정비된다면 AI 기반 음악 기술은 산업 혁신에 긍정적으로 기여할 수 있을 것이다. 이용자 역시 타인의 권리를 존중하며 기술을 책임 있게 사용하는 윤리적 태도가 중요하다.

10) KCI, "AI 커버곡의 저작권법적 쟁점과 실무적 접근 방안", 2025.

| 참고자료

- 한국저작권보호원, "AI 커버곡과 저작권 침해", 2024.07.23.,
https://www.kcopa.or.kr/lay1/bbs/S1T233C496/A/88/view.do?article_seq=5718&cpage=1&rows=10&condition=&keyword=
- 한국저작권보호원, "저작권 침해 이슈 리포트(인공지능(AI) 커버곡과 저작권 침해)", 2024.05.02.,
https://www.kcopa.or.kr/lay1/bbs/S1T283C292/F/65/view.do?article_seq=5446&cpage=1&rows=10&utm_source
- KCI, "AI 커버곡의 저작권법적 쟁점과 실무적 접근 방안", 2025.,
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART003221048>
- 한국과학기술기획평가원, "AI 커버곡, 생성 AI 기술의 새로운 시장을 형성", 2024.03.12.,
https://www.kistep.re.kr/gpsTrendView.es?mid=a30200000000&list_no=3240&nPage=1
- 한국콘텐츠진흥원, "2024 음악산업백서", 2024.09.30.,
<https://www.kocca.kr/kocca/bbs/view/B0000146/2006992.do?searchCnd=1&searchWrd=&cateTp1=&cateTp2=&useYn=&menuNo=204154&categorys=0&subcate=0&cateCode=&type=&instNo=0&questionTp=&ufSetting=&recovery=&option1=&option2=&year=&morePage=&qtp=&domainId=&sortCode=&pageIndex=2>
- phonely, "How does AI Voice Cloning Work?", 2024.12.01.,
https://www.phonely.ai/blogs/how-does-ai-voice-cloning-work?utm_source
- arxiv, "Neural Voice Cloning with a Few Samples", 2018.10.12., <https://arxiv.org/abs/1802.06006>
- arxiv, "Voice Transformer Network: Sequence-to-Sequence Voice Conversion Using Transformer with Text-to-Speech Pretraining", 2019.12.14., <https://arxiv.org/abs/1912.06813>
- arxiv, "Voice Cloning: Comprehensive Survey", 20025.05.01., <https://arxiv.org/pdf/2505.00579>
- Grand view research, "Generative AI in Music Market (2024 - 2030)"
<https://www.grandviewresearch.com/industry-analysis/generative-ai-in-music-market-report>
- KCI, "AI 커버곡의 공정 이용에 관한 연구", 2024.12.,
<https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artid=ART003146659>