

음성합성기술 기반 가창소프트웨어에 대한 인식과 개선점: 전문가 심층인터뷰 중심으로

왕 우 칭¹ · 이 철 회^{2*}¹경희대학교 포스트모던음악학과 석사과정²경희대학교 포스트모던음악학과 조교수

Recognition and Improvement of Singing Software Using Voice Synthesis Technology Based on In-Depth Interviews with Experts

Yujing Wang¹ · Chul-Hee Lee^{2*}¹Master's Course, Department of Postmodern Music, Kyung Hee University, Yongin 17104, Korea²Assistant Professor, Department of Postmodern Music, Kyung Hee University, Yongin 17104, Korea

[요 약]

본 연구는 음성합성기술 기반 가창소프트웨어가 음악 창작과 산업에 미치는 영향을 탐구하고, 개선 방향을 제시하기 위해 진행되었다. 디지털 기술의 발전과 함께 음성합성기술은 창작 효율성과 비용 절감의 가능성을 보여주었으나, 감정 표현 부족, 음색 다양성 한계, 저작권 문제와 같은 개선 과제도 확인되었다. 특히, 이러한 기술은 음악 산업에서 창작 주체성과 저작권 문제를 새롭게 제기하며, 관련 논의가 절실히 요구되고 있다. 12명의 전문가 심층 인터뷰를 통해 음성합성기술은 창작 도구로서 잠재력을 지니고 있었으나, 감정 표현 및 음질 향상, 사용자 친화적 설계, 다학문적 협력이 필요함을 확인하였다.

[Abstract]

This study was conducted to examine the influence of singing software utilizing voice synthesis technology on the process of music creation and its implications for the industry and propose potential areas for enhancement. Owing to advancements in digital technology, voice synthesis technology can improve creative efficiency and lower costs. Nevertheless, researchers have identified challenges, including limited emotional expression, lack of vocal diversity, and copyright issues. This technology raises important and expedient questions about creative independence and copyright within the music industry. Extensive interviews with 12 professionals revealed that, although voice synthesis technology has considerable potential as a creative tool, enhancements in emotional expression, sound quality, user-friendly design, and interdisciplinary collaboration are necessary to optimize its effectiveness.

색인어 : 음성합성기술, 음성합성기술 기반 가창소프트웨어, 심층 인터뷰, 음악 산업, 음악 창작**Keyword** : Voice Synthesis Technology, Singing Software Based On Voice Synthesis Technology, In-Depth Interview, Music Industry, Music Creation<http://dx.doi.org/10.9728/dcs.2025.26.1.149>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 30 November 2024; Revised 30 December 2024

Accepted 15 January 2025

*Corresponding Author; Chul-Hee Lee

Tel: +82-31-201-2863

E-mail: ch@khu.ac.kr

I. 서론

디지털화의 지속적인 발전 속에서 AI(Artificial Intelligence, 인공지능) 기술은 회화, 글쓰기, 음악 등 여러 예술분야에 적용되고 있다. 그중 음악 분야에서 활용하는 음성합성기술은 컴퓨터 소프트웨어를 사용하여 인간의 목소리를 모방하는 기술[1]로서 초기에는 음성 인식 및 음성 변환 분야에서 사용되었지만 현재는 버추얼 가수, 가이드 보컬, 기존 가수의 음성을 활용한 새로운 음악 창작 등 다양한 분야에서 널리 사용되고 있다. 보컬로이드(VOCALOID), 신시사이저 V(synthesizer V), 우타우(UTAU), 에이스 스튜디오(ACE Studio), 수노(Suno), 아이나우(Ainow)등 다양한 음성 합성 엔진 기반 가창소프트웨어의 개발 및 상용화로 사용자가 점차 늘어나고 있다. 그중 보컬로이드는 ‘하츠네 미쿠(Hatsune Miku)’, ‘SeeU’ 등 유명 버추얼 가수의 음성 합성에 사용되어 인지도가 높다. 보컬로이드는 일본의 종합 악기 및 음향 제조업체인 야마하(YAMAHA)에서 개발하였으며 보컬(Vocal)과 안드로이드(Android)의 합성어이다. 이 소프트웨어는 사용자가 가사와 멜로디를 입력하면, 사전에 저장된 라이브러리를 이용해 사람의 목소리를 인공적으로 합성하여 노래를 만들어 낸다[2]. 보컬로이드 출시는 버추얼 가수와 디지털 음악 창작의 새로운 시대를 열었고 미래의 동종 소프트웨어 발전을 위한 토대를 마련하였다.

음성 합성 엔진 기반 가창소프트웨어 기술 발전으로 인해 해당 기술의 음악 산업에 대한 영향, 개선점 및 기대 효과, 저작권 등 지적 재산권에 대한 문제가 논의되기 시작했고, 사용자 경험 및 완성도, 창작 주체성에 대한 다양한 의견이 존재한다[3]. 예를 들어, 사용자 경험 측면에서 이 기술은 비전문가의 음악 창작 활동에 크게 기여하며 음악 창작의 대중성을 높였다. 다만, 표준화된 노래 생성 방식으로 인해 창작자의 의도가 온전히 반영되기는 어렵다. 노래의 음색은 소프트웨어의 사운드 라이브러리에서 제공되며, 감정과 표현 효과는 소프트웨어에 의해 생성된다. 이로 인해 창작 주체성 측면에서 소프트웨어 제작사가 주체인지, 아니면 소프트웨어 사용자가 주체인지에 대한 논의가 대두되고 있다[4].

따라서 본 연구는 음악 분야 활동자들의 심층 인터뷰를 통해 음성합성기술 및 음성 합성 엔진 기반 가창소프트웨어 응용이 현재 음악 창작에 어떤 영향을 미치는지 탐색하였다. 개발자, 사용자, 가창자의 세 그룹으로 나누어 인터뷰를 진행했으며 각 그룹별 관점과 경험을 정리하였다. 본 연구는 구체적이고 다각화된 의견을 통해 향후 음성합성기술 및 음성 합성 엔진 기반 가창소프트웨어 발전에 일조할 수 있을 것으로 기대된다.

II. 음성합성기술(TTS, Text to Speech)

2-1 음성합성기술 발전 과정

1) 음성합성기술

음성합성기술(TTS)은 가창 합성 기술(SVS, Singing Voice Synthesis)의 기초다. 음성합성기술은 사람의 음성을 시뮬레이션할 목적으로 기계, 전자 또는 디지털 장치를 사용하여 음성을 합성하는 기술을 말한다. 음성 합성은 기계식·전자식·디지털 합성의 세 가지 발전 단계를 거쳤다.

• 기계식 합성 단계

음성합성기술의 기원은 18세기와 19세기로 거슬러 올라간다[5]. 1779년 크리스찬 크라젠슈타인(Christian Kratzenstein)은 갈대관과 서로 다른 공진기를 필터로 사용하여 인간과 유사하게 5개의 장모음을 낼 수 있는 인간의 목소리 채널 모델을 만들었다[6].

1791년 울프강 폰 캠펠렌(Wolfgang von Kempelen)은 그가 만든 “말하는 기계”(Wolfgang von Lempelen’s Speaking Machine)를 발표했다. 이 기계는 풀무로 사람의 폐를, 리드 조각으로 성대를, 호스로 성도를, 혀와 입술도 시뮬레이션했다. 기계의 ‘혀’와 ‘입술’을 제어함으로써 창작자는 기계가 하나의 소리를 내거나 소리 조합을 만들 수 있었다[7].

• 전자식 합성 단계

19세기 말과 20세기 초에 개발된 전자 기술은 음성합성기술의 발전을 크게 촉진했으며 당시의 전자 기술은 전자 음성 합성기를 탄생시켰을 뿐만 아니라 음성합성기술 연구에도 사용되었다.

벨 연구소(Bell Labs)는 1930년대에 보코더(Vocoder)를 발명했다. 음성을 자동으로 음조와 공명으로 분해했다. 이 기술은 1939년 호머 더들리(Homer Dudley)에 의해 키보드 신시사이저인 VODER(Voice Operating Demonstrator)로 개선되었으며, 이는 음성 신시사이저로 알려진 최초의 장치이다[8].

• 디지털 합성 단계

소리를 연구하는 과정에서 연구자들은 발성을 시뮬레이션하기 위해 포르مان트 합성(Formant Synthesis) 방법을 만들었다. 포르مان트의 매개변수는 사람의 언어로 구성된 말의 제현 원리를 사용하여 합성하는 방식이다. 첫 번째 포르مان트 합성기는 1953년 월터 로렌스(Walter Lawrence)에 의해 제안된 3개의 병렬 전자 공진기로 구성되었다[9]. 그 이후로 음성합성기술은 점차 포르مان트 합성 기술이 주도하는 단계에 진입했다.

컴퓨터 기술의 급속한 발전과 함께 음성합성기술은 점차 디지털 합성 단계에 진입하여 포르مان트 합성 기술을 컴퓨터 기술과 통합하고 디지털로 표시할 수 있었다. 컴퓨터 기술, 특히 음성 부호화 및 디코딩 기술의 발달로 음성합성기술은 파

형 접합법(Concatenative Synthesis)을 도입했다. 파형 접합법의 원리는 준비된 사운드 라이브러리에서 원하는 음성 단위를 선택하여 텍스트 기반 언어 운율 분석 결과에 따라 접합하여 음성을 생성하는 것이다. 1980년대 말에 제안된 기본 음성 동기 중첩(Pitch Synchronous Overlap and Add) 알고리즘은 음성 스플라이싱 합성의 일부 문제를 잘 해결했으며 파형 스플라이싱에서 가장 일반적인 알고리즘이다. 음성 동기 중첩 알고리즘에 의해 합성된 소리는 원래의 주요 음질 특성을 유지할 수 있을 뿐만 아니라 높은 선명도와 자연스러움을 가질 수 있어 음성 합성 품질을 크게 향상시킬 수 있었다[10].

20세기 말 연구자들은 훈련 가능한 음성 합성(Trainable TTS) 기술에 주목했다. 기본 원리는 음성 라이브러리에서 추출된 음성 단위의 특성 매개변수를 얻은 뒤 이를 모델링하고 알고리즘을 통해 매개변수 또는 모델을 훈련한 결과를 사용하여 음성 합성 시스템을 구축하는 것이다[11].

2) 가창 음성합성기술(SVS - Singing Voice Synthesis)

노래와 음성은 사용되는 발성기관과 발성 원리가 유사하지만, 유성음과 무성음의 비율에 큰 차이가 있다. 노래에서는 약 90%가 유성음, 나머지 10%가 무성음인 반면, 영어를 기준으로 음성에는 약 60%가 유성음, 나머지 40%가 무성음[12]. 노래는 동일한 선율이라도 다양한 발성 기법에 따라 결과물이 상이할 수 있다. 또한 같은 가사라도 선율의 길이나 높이에 영향으로 서로 다른 음성적 특징이 발현된다. 따라서 일반적인 음성 합성 모델을 가창 음성 합성 모델로 직접 전환하는 것은 바람직하지 않다. 가창 음성합성기술의 발전은 음악 합성 기술과 음성합성기술의 발전을 기반으로 하며, 음정과 리듬의 제어가 핵심 기술이다. 특히 비브라토(vibrato)는 노래의 감정 표현에 주요한 요소임으로 가창 음성 합성의 중심 연구 내용이다.

최초의 관련 연구는 1950년대 미국 벨 연구소에서 이루어진 일련의 전자 음악 실험이다. 벨 연구소에 이어 맥스 매튜스(Max Mathews)는 켈리, 로크바움과 함께 초기 가창 음성 합성기술을 연구했다[13]. 또 다른 초기 연구는 VOCODER 음성 생성기 중 하나인 초기 음성합성기술이다. 1986년에 맥 컬리(McAulay)와 콰티에리(Quatieri)는 사이누소이달 음성 모델(Sinusoidal Speech Modeling)의 가창 음성합성기술을 제안했다. 이 모델은 화음, 시간 스케일링 및 음조를 분리하여 음정 등의 매개변수를 조정할 수 있으며, 정확한 합성 결과를 얻을 수 있었다[14]. 사이누소이달 음성 모델의 결과물은 완성도 면에서는 뛰어났으나, 수집해야 하는 매개변수가 많고 계산 과정이 번거롭기 때문에 연구자들은 파형 접합 기술(Waveform concatenation technology)에 주목하게 되었다. 파형 접합 기술은 합성된 음성의 자연스러움은 초기의 음성합성기술을 대폭 상회하였으며, 실제 인간의 음성과도 구별하기 어려운 결과물을 제작했다. 다만, 인간이 표현하는 미세한 비브라토와 음 이동 시 자연스러운 전환 등 일부 요소에

서는 다소 부족했다. 도쿄 대학교의 메론 Y(Meron Y)는 박사는 대규모 데이터베이스를 사용하여 가창 음성 합성의 ‘인간성’을 향상시키기 위해 파형 접합을 기반으로 하는 가창 음성 합성 시스템을 제안했다. 이 시스템은 가수들의 노래 정보를 분석하여 다양한 스타일, 비브라토 및 음정 간 전환과 같은 특성에 기반한 수학적 모델을 얻었다[10].

2-2 음성합성기술 기반 가창소프트웨어의 개념과 정의

음성합성기술 기반 가창소프트웨어는 현대 음성합성기술을 활용하여 사람의 목소리를 모방하고 자동으로 음악을 생성하는 소프트웨어 도구이다. 이 소프트웨어는 주로 스코어 에디터(Score Editor), 싱어 라이브러리(Singer Library), 그리고 합성 엔진(Synthesis Engine)으로 구성된다. 스코어 에디터는 피아노 롤(piano-roll) 형식의 편집기로, 사용자가 음표, 가사, 그리고 미디 컨트롤 익스프레션(Expression)을 입력하는 데 사용된다. 싱어 라이브러리는 다양한 보컬 오디오 샘플링 음원을 포함하고 있으며, 해당 언어의 발음 및 표현의 다양한 조합을 담고 있다. 마지막으로 합성 엔진은 가창 음원 생성의 중심 부분으로, 악보 정보를 기반으로 선택한 음원의 높낮이와 음색을 조절하고 이를 결합하여 필요한 가창 음원을 생성한다[15]. 이러한 소프트웨어의 핵심은 내장된 음성 합성 엔진에 있다. 이 엔진은 입력된 텍스트 정보를 분석하고 이해하여 자연스럽게 부드러운 음성으로 변환하는 데 인공지능 기술을 활용한다. 음성 합성 엔진은 다량의 음성 데이터를 학습하고 훈련하여 음높이, 음색, 억양 등을 포함한 인간의 발화 특성을 모방하며, 이를 통해 고품질의 음성 합성을 실현할 수 있다. 음성 합성 엔진(기술) 기반 가창소프트웨어는 노래 합성 및 가창 측면에서 독특한 특징과 기능을 가지고 있다. 입력된 가사와 악보를 기반으로 노래의 음악 콘텐츠를 자동으로 생성할 뿐만 아니라 사용자가 설정한 음색과 감정 스타일에 따라 원하는 노래 효과를 자동으로 불러줄 수 있다. 사용자는 간단한 작업을 통해 적절한 노래 스타일, 음색 설정 및 감정 표현을 선택하여 개인화된 노래 작품을 빠르게 생성할 수 있다.

음성 합성 엔진(기술) 기반 가창소프트웨어는 음성합성기술을 사용한 혁신적인 도구이다. 사용자에게 고품질 곡 합성 및 가창 서비스를 제공할 뿐만 아니라 음악 창작의 편의성을 촉진하여 더 많은 사람들이 음악 창작 과정에 참여할 수 있도록 하여 음악 산업의 다양성과 지속가능성을 부여한다.

III. 심층 인터뷰에 대한 연구방법 및 연구문제

3-1 연구방법

본 연구는 선행연구[16]-[18]를 참고하여 음악 분야 활동자들이 음성합성기술 및 음성 합성 엔진 기반 가창소프트

웨어가 음악 산업에 미치는 영향, 기술의 개선점 및 기대 효과, 저작권등 지적 재산권에 대한 인식, 개발자, 사용자, 가창자 간 입장을 밝히는데 목적이 있다. 기존 연구는 주로 순수 기술 연구나 버추얼 가수와 같은 파생상품이 주를 이뤘다. 그러나 음악 분야에서 활동자들의 음성합성기술에 대한 인식과 응용에 대한 연구는 없었기 때문에 본 연구는 탐색적 연구로서 심층 인터뷰(in-depth interview)를 통한 질적 연구방법이 사용되었다.

본 연구는 표 1과 같이 음악 분야 활동자 12명을 대상으로 심층 인터뷰를 진행하였다. 음성합성기술 및 음성 합성 엔진 기반 가상소프트웨어를 대상으로한 연구인 만큼 12명의 인터뷰 대상은 음악 소프트웨어 개발자, 사용자, 가창자로 나누어 분류하였다. 또한, 연구의 포괄성을 위해 위에 언급한 공통적인 질문 외에 다른 입장에 대한 질문을 별도로 설계하였다. 인터뷰는 구조화된 질문지를 가지고 1인당 약 30분에서 60분 정도 진행되었다. 인터뷰 내용의 진실성과 정확성을 위해 참여자들에게 허락을 받은 후에 녹음했고, 인터뷰 끝난 후에 녹음 내용을 정리하였다. 마지막으로, 서로 다른 입장에 있어서 공통된 의견과 다른 의견을 클러스터링하고 연구 결과를 제시하였다.

표 1. 인터뷰 참여자 인구통계 정보

Table 1. Interviewee demographic characteristics

Category	Number	Years of Working	Age	Gender
Developer	A	5	31	Male
	B	4	27	Male
	C	10	35	Male
	D	3	28	Female
User	E	5	27	Female
	F	3	27	Male
	G	6	29	Female
	H	3	25	Female
Competitor	I	7	27	Female
	J	6	29	Male
	K	6	27	Female
	L	4	28	Female

3-2 연구문제

표 2는 본 심층 인터뷰에서 공통 질문을 제시한다. 공통 질문은 모든 인터뷰 대상자에게 동일하게 제시된 질문들로, 음성합성기술의 전반적인 영향을 파악하고자 하는 목적을 가진다. 이를 통해 기술이 음악 산업에 미치는 영향을 광범위하게 분석할 수 있으며, 창작 활동에서 음성합성기술의 장점과 한계를 이해하는 데 중요한 기초 자료를 제공한다. 또한, 그림 1에 제시된 데이터[19]와 음성합성기술이 실제 가수를 대체할 수 있는 가능성에 대한 의견을 통해 음성합성기술 기반 가상 소프트웨어를 포함한 생성형 AI 기술이 미래의 음악 산업에

표 2. 공통질문

Table 2. The same question

Category	Question	Number
Impact on the Music Industry	What do you think is the impact of the development of speech synthesis technology on the music industry as a whole?	1-1
	What do you think are the advantages of speech synthesis technology to creative activities?	1-2
	Do you think speech synthesis technology can completely replace real singers?	1-3
Technology Improvements and Expected Effects	What do you think are the improvements in speech synthesis technology?	1-4
	What kind of effort and support do you think is needed to improve speech synthesis technology?	1-5
	What expectations do you have for speech synthesis technology?	1-6
The recognition of intellectual property rights such as copyrights	Where do you think speech synthesis technology can cause copyright and legal issues?	1-7
	What action do you think is necessary to strengthen copyright protection in speech synthesis technology?	1-8
	What kind of effort do you think is needed to help the public better understand copyright and intellectual property issues for speech synthesis technology?	1-9

서 차지할 위치를 예측할 수 있다.

표 3은 본 심층 인터뷰에서 개발자 대상 질문들을 제시한다. 개발자 대상 질문은 음성합성기술의 기술적 발전과 그 한계를 가장 잘 이해하는 전문가들에게 제시된 질문들이다. 이 질문들은 음성합성기술의 발전 과정, 현재의 한계, 그리고 감정 표현 및 자연스러운 발음을 구현하기 위한 기술적 난제를 파악하는 데 목적이 있다. 또한, 음성합성기술이 향후 어떻게 발전할 수 있을지에 대한 개발자들의 비전과 노력들을 들여다보려는 목적을 가진다.

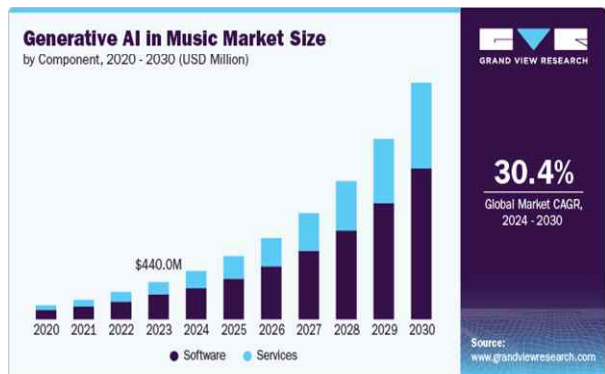


그림 1. 음악시장에서 생성형 AI의 시장 규모 현황과 전망

Fig. 1. The state of the market for generative AI in the music industry

표 3. 개발자 대상 질문

Table 3. Question for developers

Question	Number
What do you think is the biggest advancement in speech synthesis technology in the past few years?	2-1
What is important for speech synthesis techniques that are not mechanical and natural (so hard to distinguish from humans)?	2-2
What do current speech synthesis technologies lack in terms of emotional expression?	2-3
How do I guarantee the copyright and intellectual property protection of speech synthesis technology?	2-4
What is the most technical difficulty when developing speech synthesis software?	2-5
Have you ever collaborated with experts in other fields during the development process?	2-6

표 4. 사용자 대상 질문

Table 4. Question for user

Question	Number
How did you come across speech synthesis software?	3-1
Have you tried speech synthesis software from other brands? What are the pros and cons of each software?	3-2
Do you think the tone selection of these speech synthesis software is sufficiently diverse?	3-3
What do you think are the main strengths of these speech synthesis software?	3-4
Have you used a combination of speech synthesis and actual singing?	3-5
Do you think speech synthesis singing software is easy and convenient to use?	3-6

표 4는 본 심층 인터뷰에서 사용자 대상 질문들을 제시한다. 사용자 대상 질문은 실제 음성 합성 소프트웨어를 사용하는 사람들이 경험하는 문제와 요구사항을 이해하기 위해 제시된 질문들이다. 이 질문들은 소프트웨어의 사용 용이성, 음성합성기술에 대한 기대, 그리고 기술 사용에서의 장단점 등을 파악하는 데 목적이 있다. 사용자의 피드백을 통해 소프트웨어 개선점과 사용자 경험을 향상시킬 수 있는 방향에 대한 정보를 얻을 수 있다.

표 5는 본 심층 인터뷰에서 가창자 대상 질문들을 제시한다. 가창자 대상 질문은 음성합성기술이 실제로 노래를 부르는 사람들에게 미치는 영향을 파악하기 위해 제시된 질문들이다. 이 질문들은 음성합성기술이 가창자의 직업에 미치는 영향과, 그들이 느끼는 감정 표현 능력이나 노래 기교에서의 부족한 부분을 파악하는 데 목적이 있다. 또한, 가창자가 음성합성기술과의 협업 가능성이나 활용 경험에 대해 어떻게 생각하는지에 대해 알아보기 위한 질문이다.

표 5. 가창자 대상 질문

Table 5. Question for singer

Question	Number
What do you think about making music with speech synthesis technology?	4-1
What do you think about the ability of speech synthesis technology to express emotions?	4-2
What do you think speech synthesis technology lacks in singing techniques?	4-3
How do you think speech synthesis technology affects your job?	4-4
Have you considered collaborating with a speech synthesis technology developer?	4-5
Have you ever thought about recording your voice through speech synthesis technology and using it to create music?	4-6

IV. 공통질문에 대한 연구결과

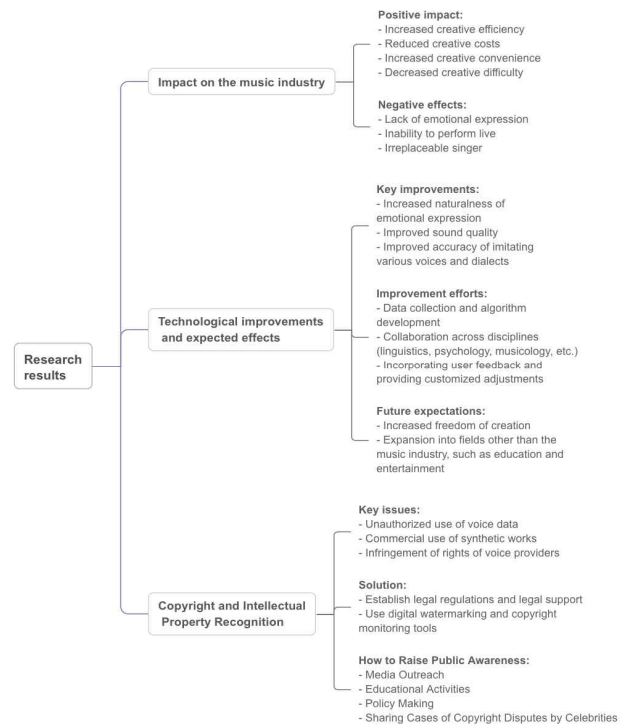


그림 2. 공통질문에 대한 연구결과 정리

Fig. 2. Summary of research results on common questions

4-1 음악 산업에 미친 영향

본 연구 결과는 표 6에서 보여지듯이 음악 분야 활동자들은 음성합성기술이 음악 산업에 미치는 영향에 대해 대체로 긍정적인 태도를 보였다. 장점으로는 창작 효율성 향상, 창작 비용 절감, 창작의 편의성 증가, 창작 난이도 감소 등이 있다. 그러나 대부분의 참여자들은 현재 또는 미래에 음성합성기술

이 실제 가수를 대체할 수 없다고 생각하며, 그 주된 이유로 감정 표현의 부족과 라이브 공연의 불가능함을 꼽았다.

인터뷰 응답자들에 따르면, 음성합성기술은 가수의 녹음 비용을 절감하고 창작 비용을 낮추며, 음악 제작의 효율성을 높이는 데 큰 기여를 하고 있다(A, D, E). 특히, 이 기술은 다양한 음색을 쉽게 찾을 수 있는 기능을 통해 음악가들이 새로운 창작을 시도하는 데 도움을 주고 있다(A). 또한, 전통적인 가수과 버추얼 가수가 공존함으로써 새로운 시장 구조가 형성될 가능성이 있으며, 이는 음악 산업의 패러다임 전환을 야기할 수 있다(A, K).

하지만 일부 응답자들은 음성합성기술이 현재 주로 개인적 오락이나 간단한 배경음악(BGM, Back Ground Music)과 같은 제한된 영역에 적용되고 있으며, 전문 음악 제작에서는 아직 큰 영향을 미치지 못하고 있다고 언급하였다(B, C, F). 또한, 이 기술이 음악 창작의 문턱을 낮추고 효율성을 높였지만, 작품의 예술성이나 감정 표현에서 한계가 있을 수 있다는 우려도 있었다(G).

표 6. 음악 산업에 미친 영향에 대한 결과 정리
Table 6. Summary of the impact on the music industry

Question Number	Key Responses
1-1	Cost Reduction in Creation, Provision of Diverse Vocal Tones, Coexistence of Traditional and Virtual Singers, Creation of New Markets, Application in Limited Areas
1-2	Improved Creative Freedom and Efficiency, High-Quality Music Creation Accessible to Non-Professionals, Elimination of Language Barriers, Opportunities for Experimenting with Diverse Vocal Tones, Economical Alternative
1-3	Significant Differences from Human Singers in Emotional Expression and Unique Vocal Style, Difficult to Replace Emotional Expression and Interaction with the Audience in Live Performances, Limited Replacement Potential

4-2 기술의 개선점 및 기대 효과

본 연구 결과는 표 7에서 보여지듯이 음악 분야 활동자들은 음성합성기술의 개선점 및 기대 효과에 대해 주로 자연스러움과 현실감을 높여야 한다는 의견이 주를 이루는 것으로 나타났다. 이 기술을 개선하기 위해 필요한 노력과 지원에 대해 대부분의 참여자들은 기술 발전, 데이터 수집, 기관 협력, 그리고 사용자 피드백의 중요성을 언급했다. 또한, 참여자들은 모두 이 기술이 미래에 실제 가수의 수준에 도달할 수 있기를 기대하고 있다.

본 연구의 결과, 인터뷰 응답자들에 따르면 음성합성기술은 감정 표현의 자연스러움을 높이고, 음질의 개선이 필요하다는 점이 주요한 개선점으로 언급되었다(A, B, D, H, I). 이 기술의 감정 표현 능력이 향상되면, 실제 가수의 수준에 도달할 수 있을 것이라는 기대가 있었다(A, C, K). 또한, 다양한

음색과 사투리의 모방이 정확해져야 한다는 의견도 제시되었으며, 이를 위해 더 많은 데이터와 알고리즘의 발전이 필요하다고 강조되었다(D, F, H).

개선 노력을 위해서는 개발자들의 지속적인 연구와 더불어 다양한 학문 분야 간의 협력이 필수적이라는 점이 강조되었다(A, D, F). 특히, 언어학, 심리학, 음악학 등 여러 학문 간의 협력을 통해 감정 표현과 음색의 다양성을 높이는 것이 필요하다는 의견이 있었다(D, F). 사용자 피드백을 반영하여 기술을 개선하고, 이를 통해 사용자 맞춤형 조정 기능을 제공하는 것도 중요한 개선 방안으로 제시되었다(I, L).

향후 기대되는 효과로는 음성합성기술이 더욱 인간과 유사한 감정 표현을 가능하게 하여, 창작자들이 자유롭게 창작물을 조정할 수 있는 능력을 가지게 될 것이라는 점이었다(A, B, G). 또한, 음악 창작뿐만 아니라 교육 등 다양한 분야에서의 활용 가능성이 제기되었으며, 이를 통해 음악 산업의 새로운 가능성을 열어줄 수 있을 것이라는 기대가 있었다(F, L).

표 7. 기술의 개선점 및 기대 효과에 대한 결과 정리
Table 7. Summary of results on technological improvements and expected effects

Question Number	Key Responses
1-4	Enhancement of Naturalness, Improvement in Emotional Expression Ability, Need for Tone Stability and Smoothness, Addition of Precise Voice Control and Vocal Variety, Improvement in Breathing and Pronunciation, Need for Mimicking Various Voices and Dialects
1-5	Developer Collaboration and Data Accumulation, Research to Strengthen Model Flexibility, Collection of Diverse Data and Interdisciplinary Collaboration, Support in Human Resources and Technical Efforts, Incorporation of User Feedback and System Improvements
1-6	Significant Differences from Human Singers in Emotional Expression and Unique Vocal Style, Challenging to Replace Emotional Expression and Interaction with the Audience in Live Performances, Limited Potential for Replacement

4-3 저작권등 지적 재산권에 대한 인식

본 연구 결과는 표 8에서 보여지듯이 음악 분야 활동자들은 음성합성기술 사용 시 발생할 수 있는 저작권 및 법적 문제로 주로 음성 데이터의 저작권 허가와 작품의 사용 목적 두 가지에 중점을 두고 있다. 이와 관련하여, 참여자들은 관련 법률 및 규정의 제정이 매우 중요하고 필수적이라고 공통적으로 언급했다. 또한, 음성합성기술에 대한 저작권 및 지식 재산권 문제에 대한 대중의 인식을 높이기 위한 방법으로 미디어 홍보와 관련 사건의 보도가 가장 효과적이라고 생각하는 경향이 강했다.

본 연구의 결과, 인터뷰 응답자들에 따르면 음성합성기술의 사용 중 저작권 문제가 발생할 수 있는 주요 부분으로는

음성 데이터의 무단 사용, 합성된 작품의 상업적 사용, 그리고 음성 제공자의 권익 침해 등이 있었다(A, B, D, E, G, H, I, J, L). 특히, 합성된 작품이 상업적으로 사용될 때 원저작자의 허가를 받지 않으면 저작권 침해 문제가 발생할 수 있다는 점이 강조되었다(A, B, D, J).

저작권 보호를 강화하기 위해 필요한 조치로는 법적 규제의 마련과 법적 지원이 중요하다는 의견이 있었다(A, B, C, G, H, I, K, L). 법적 보호가 명확하게 이루어져야만 음성 데이터 제공자와 창작자의 권익을 보호할 수 있으며, 이를 위해 업계 전반에서 절차와 규정을 통일하는 것이 필요하다는 제안이 있었다(A, B, H). 또한, 디지털 워터마크나 저작권 모니터링 도구 등을 활용하여 저작권 보호를 강화할 수 있다는 기술적 조치도 제시되었다(C, F).

대중의 저작권 인식을 높이기 위한 노력으로는 미디어 홍보, 교육 활동, 정책 제정 등의 방법이 필요하다는 의견이 있었다(A, B, D, E, G, H, J, K, L). 특히, 구체적인 사례를 공유하거나 유명 인물의 저작권 분쟁 사건을 알리는 것이 대중의 관심을 끌고 저작권 문제에 대한 이해를 높이는 데 효과적일 것이라는 제안이 있었다(A, B, I, J).

표 8. 저작권 등 지적 재산권에 대한 인식에 대한 결과 정리

Table 8. Summary of results on perceptions of copyright and intellectual property rights

Question Number	Key Responses
1-7	Issues with Voice Data Usage Permissions, Copyright Issues for Commercial Use, Unauthorized Use of Specific Singers' Voices
1-8	Establishment of Legal Regulations, Resolution of Clear Copyright Attribution Issues, Development of Digital Watermarks and Monitoring Tools, Standardization of Procedures and Regulations
1-9	Media Promotion and Educational Activities, Sharing of Related Cases, Promotion by Official Institutions, Publicizing Incidents Involving Famous Individuals

V. 개발자, 사용자, 가상자 간 입장에 대한 연구결과

5-1 개발자 입장

본 연구 결과는 표 9에서 보여지듯이 인터뷰에 응답한 개발자들은 음성합성기술의 주요 개선점으로 자연스러운 감정 표현 및 음질 향상을 꼽았다. 현재 기술 수준에서 세밀한 감정 변화를 표현하는 데 한계가 있으며, 창작자가 감정을 주관적으로 제어하는 데 있어 추가적인 기술적 지원이 필요하다는 점이 지적되었다. 이 외에도 다학문적 협력의 중요성을 강조하며, 향후 더 자연스러운 합성 음성을 구현하기 위해서는 음악, 언어, 법무 전문가의 협력이 필수적임을 언급했다.

인터뷰 응답자들에 따르면, 음성합성기술은 특히 딥러닝과 데이터 축적의 발전으로 인간의 음성과 유사한 수준의 자연

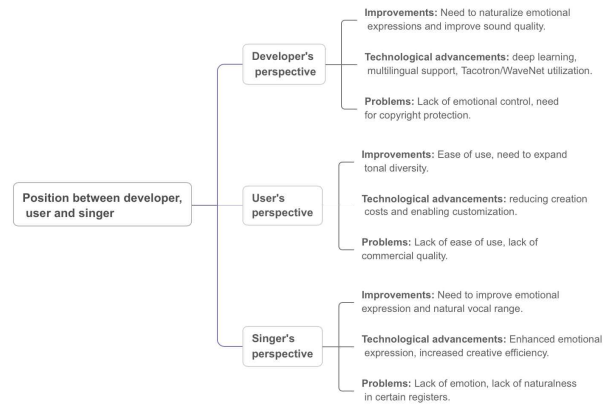


그림 3. 개발자, 사용자, 가상자 간 입장에 대한 연구결과 정리
Fig. 3. Summary of research findings on the positions of developers, users, and singers

표 9. 개발자 대상 질문에 대한 결과 정리

Table 9. Summary of results from questions targeted at developers

Question Number	Key Responses
2-1	Deep Learning Enhancement, Data Accumulation, Improvement of Naturalness, Model Integration
2-2	Sound Quality, Control of Dynamics, Enhanced Emotional Expression, Tone Stability, Pitch Maintenance, Singing Techniques, Language Diversity, Control of Tone and Speed
2-3	Lack of Subtle Variation in Emotional Expression, Difficulty in Creator Control, Limitations in Emotional Transformation
2-4	Securing Legal Authority, Clarification of Copyright, Digital Watermarking, Data Usage Permissions, Prevention of Unauthorized Use
2-5	High-Quality Audio Generation, Realistic Voice Implementation, Challenges in Model Architecture Design, Lack of Multilingual Data, Natural Matching of Emotion and Lyrics, Need for Large-Scale Data
2-6	Music Experts, Language Experts, Legal Experts, Need for Multidisciplinary Collaboration

스러운 음성을 구현하는 데 큰 기여를 하고 있다(A, B, D). 이 기술은 현재 다국어 지원을 가능하게 하며, Tacotron 및 WaveNet과 같은 모델을 통해 더 효율적으로

자연스러운 음성을 생성할 수 있다(A, D). 또한, 데이터 주석 및 정제 비용을 절감하여 개발 비용을 낮추고, 다양한 응용 분야에서 음성 합성의 가능성을 확대하고 있다(B).

하지만 일부 응답자들은 여전히 감정 표현에서 부족함을 지적하며, 생성된 음성에서 감정 강약 변화 및 세밀한 감정 조절이 어려워 실제 사람의 감정을 완벽히 모방하는 데 한계가 있다고 언급하였다(A, B, C, D). 현재의 기술은 주로 기본적인 감정 표현에 머무르고 있으며, 창작자가 감정의 세부적 조절을 수행하는 데 한계가 있다는 우려가 있었다(B, C). 또한, 저작권 보호와 관련하여 데이터 사용 권한의 명확화와 디지털 워터마크와 같은 기술적 보호 장치의 필요성도 강조되

었다 (A, C, D).

5-2 사용자 입장

본 연구 결과는 표 10에서 보여지듯이 인터뷰에 응답한 사용자들은 음성 합성 소프트웨어의 주요 개선점으로 사용 편의성과 음색 다양성 확대를 꼽았다. 응답자들은 현재 대부분의 소프트웨어가 복잡한 인터페이스와 제한된 작업 흐름으로 인해 일반 사용자에게는 접근성이 낮고, 전문 작업에서도 불편함이 있다고 평가하였다. 특히, 기존 작업 파일을 소프트웨어에 통합하거나 외부 음원을 불러오는 기능이 부족하다는 점이 지적되었다. 또한, 음색 선택의 다양성에 대해서는 상반된 의견이 있었다. 일부 응답자들은 충분히 다양한 음색을 제공하는 소프트웨어가 존재한다고 평가한 반면, 특정 독특한 음색이나 맞춤형 음색을 훈련할 수 있는 기능이 부족하다는 지적도 있었다. 음성 합성을 통해 새로운 음색을 만들어내는 기능은 사용자의 창작 욕구를 충족시킬 수 있는 중요한 요소로 언급되었다.

인터뷰 응답자들은 음성 합성 소프트웨어가 창작 과정에서의 비용 절감과 작업 효율성 향상에 중요한 역할을 한다고 언급하였다(E, F, H). 특히, 초보자도 쉽게 사용할 수 있는 인터페이스와 사용자 정의 음색 훈련 가능성은 기술의 주요 강점으로 평가되었다(G, H).

그러나 대부분의 응답자들은 소프트웨어의 사용 편의성과 음색 다양성이 부족하며, 상업적 용도로는 품질 및 자연스러움에서 미흡하다는 점을 지적하였다(E, F, G). 이와 함께, 기술이 아직 발전 초기 단계에 있으며, 사용자 요구를 충족하기 위해 지속적인 개선이 필요하다는 점도 강조되었다(H).

표 10. 사용자 대상 질문에 대한 결과 정리

Table 10. Summary of results from questions targeted at users

Question Number	Key Responses
3-1	Curiosity, Work Requirements, Company Guidelines, Experimentation and Use of Various Software
3-2	Functional Differentiation, Lack of User-Friendliness, Differences in Vocal Variety, Presence of Software Specialized in Voice Separation and Specific Features, Capability for Tone and Lyric Editing
3-3	Lack of Specific Vocal Tones, Need to Complement Diversity, Requirement for Voice Package Purchases
3-4	Reduction in Creation Costs, Improved Work Efficiency, Easy Creation for Beginners, Availability of Various Vocal Options
3-5	Non-usage, Difficulty in Combined Use, Opinion that Independent Use is Sufficient
3-6	Generally Considered Difficult, Suitable for General Users, Insufficient for Commercial Use, Primarily Used for Technical Experimentation, Requires Model Training

5-3 가창자 입장

본 연구 결과는 표 11에서 보여지듯이 인터뷰에 응답한 가창자들은 음성합성기술 기반 가창소프트웨어의 가장 큰 한계로 일부 소프트웨어의 감정 표현 능력의 부족을 지적하였다. 이러한 소프트웨어는 현재 기술 수준에서 단조롭고 평면적인 표현에 그치며, 고음과 낮은 음역 등 특정 음역에서의 자연스러움도 부족하다는 평가가 많았다. 응답자들은 이러한 소프트웨어가 음악 창작에서 규격화되고 접근성을 높이는 데 기여할 수 있지만, 프로페셔널 제작에서는 여전히 한계가 있다고 평가하였다. 특히, 맞춤형 음색 개발에 관심을 보인 응답자들은 개발자와의 협력을 통해 창작자 자신의 음색을 기반으로 한 새로운 도구를 개발할 가능성을 제시하였다.

한편, 기술 발전과 더불어 감정 표현 및 기교 강화, 사용성 개선이 이루어진다면, 음성합성기술이 창작 과정의 효율성을 높이고, 새로운 협업 및 창작의 기회를 제공할 수 있을 것으로 기대된다. 이러한 연구 결과는 음성합성기술이 음악 창작의 보조적 도구에서 벗어나, 창작과 표현의 핵심적인 기술로 자리 잡는 데 기여할 가능성을 시사한다.

또한, 인터뷰 응답자들은 음성합성기술이 음악 창작에서 새로운 가능성을 열어줄 수 있다고 평가하였다(I, J, K, L). 특히, 창작 비용 절감, 창작 과정의 효율성 증대, 대체 가창 도구로서의 가능성은 이 기술의 주요 강점으로 언급되었다(I, J, L). 또한, 맞춤형 음색 개발 및 감정 표현 개선과 같은 추가적인 기능 구현이 기술 발전의 핵심 요소로 평가되었다(K).

표 11. 가창자 대상 질문에 대한 결과 정리

Table 11. Summary of results from questions targeted at singers

Question Number	Key Responses
4-1	Standardized Creative Support, Limitations in Professional Production, Lack of Natural Results, Suitable for Demos and Entertainment Purposes, Expanded Public Accessibility
4-2	Lack of Emotional Expression, Monotonous and Limited Expression, Merely Imitation of Emotions
4-3	Lack of Naturalness in High/Low Pitch Ranges, Absence of Spatial Depth and Flat Sound, Monotonous Expression Due to Lack of Emotion, Absence of Special Vocal Techniques
4-4	No Direct Impact on Field Due to Technology, Potential for Auxiliary Use of Technology, Opportunities for Collaboration and Innovation
4-5	Collaboration Not Considered, Interest in Collaborative Development of Customized Vocal Tones, Previous Experience with Collaboration
4-6	Positive Attitude Towards Recording One's Own Voice, Skeptical About Its Use in Creation, Interested in Experimentation and Learning, Expectation and Curiosity About Voice Recording

그러나 인터뷰 응답자들은 감정 표현의 부족, 단조로운 기교, 특정 음역에서의 자연스러움 부족 등 현재 기술의 한계에 대해 지적하였다(I, J, K, L). 특히, 상업적 음악 제작 및 라이브 공연에서는 여전히 사용하기 어렵다는 점에서 기술의 한계가 명확하다고 평가하였다(I, J, L). 이에 따라 감정 표현 강화, 기술의 자연스러움 향상, 사용 편의성 개선이 필요하다는 점이 강조되었다(J, K, L).

이와 함께, 음성합성기술이 창작 과정의 보조적 도구로 자리 잡기 위해서는 기술적 품질 개선뿐만 아니라, 창작자의 요구에 부합하는 맞춤형 솔루션 제공과 개발자와의 협업 기회 확대가 필요하다는 의견이 제시되었다(K, L).

VI. 결론 및 논의

6-1 연구 요약 및 의의

본 논문은 음성합성기술 기반 가창소프트웨어를 연구 대상으로 삼고, 전문가 심층 인터뷰 분석을 중심으로 논의하였다. 연구의 목적은 음악 분야의 다양한 활동자에 대한 심층 인터뷰를 통해 음성합성기술과 소프트웨어가 현재 음악 창작에 미치는 영향을 연구하여 향후 이 기술이 음악 창작에 어떻게 적용되고 발전할 수 있을지에 대한 보다 포괄적인 이해와 지침을 제시하고, 음악 산업의 종사자들에게 유용한 통찰에 대한 제안, 미래의 연구와 실천에 방향성을 새로운 관점으로 연구하는 초석이 되는 가치가 있다.

심층 인터뷰 질문과 응답자의 피드백을 바탕으로, 다음과 같은 결론을 도출할 수 있다.

첫 번째로, 음성합성기술은 음악 창작의 효율성을 높이고 비용 절감을 가능하게 하며, 전통적인 녹음 방식을 대체할 수 있는 잠재력을 가지고 있다. 그러나 감정 표현과 예술성에서 여전히 한계를 보이며, 실제 가수의 감정 표현과 예술적 깊이를 완전히 대체하기에는 부족함이 있다.

두 번째로, 현재 음성합성기술은 감정 표현과 음질 자연스러움에서 큰 발전 가능성을 가지고 있다. 더 나은 음성 합성을 구현하기 위해서는 데이터 훈련 범위의 확장과 사용자 인터페이스의 개선이 필요하다. 이러한 발전을 통해 비전문가들도 더 쉽게 창작 도구를 활용할 수 있게 될 것이다.

세 번째로, 음성합성기술의 발전은 새로운 저작권 및 지식 재산권 문제를 야기할 수 있다. 특히 음성 라이브러리의 사용권한과 합성 음성이 특정 가수의 목소리와 유사할 경우 발생할 수 있는 법적 분쟁에 대한 우려가 존재한다. 이를 해결하기 위해서는 효과적인 저작권 보호 체계의 구축과 관련 법률에 대한 대중적인 교육이 필요하다.

마지막으로, 본 연구는 음성합성기술에 대한 저작권 보호와 법적 대응을 위한 정책적 제안을 제시한다. 현재 한국과 중국의 저작권법은 이 기술의 일부 측면을 다루고 있지만, 보다 구체적이고 명확한 가이드라인이 필요하다. 합성 음성의

저작권 문제를 해결하기 위한 기술적 추적 시스템 구축과 대중에 대한 법적 교육을 통해, 기술 발전과 법적 보호를 동시에 이루어낼 수 있을 것이다. 향후 저작권법의 개선과 함께 기술 발전을 지원하는 정책적 노력이 필요하다.

음성합성기술은 음악 산업에 혁신적인 기회와 도전을 가져왔다. 현재 기술에는 개선할 부분이 많지만, 창작 도구로서의 잠재력은 무시할 수 없다. 향후 기술의 지속적인 최적화와 저작권 문제에 대한 효과적인 대응을 통해 음성합성기술의 음악 창작 분야 응용 전망은 더욱 밝아질 것이다.

6-2 향후 계획

본 연구는 음성합성기술의 발전과 음악 창작에 미치는 영향을 심도 깊게 다루었으며, 향후 연구에서는 그 결과를 더욱 확장하고 심화시키기 위한 구체적인 계획을 마련하고자 한다. 이를 통해 연구의 범위와 깊이를 더욱 넓히고, 향후 이 기술이 음악 산업과 저작권 분야에서 어떻게 발전할 수 있을지에 대한 중요한 통찰을 제공할 수 있을 것이다.

첫째, 기술적 발전을 위한 연구에 집중할 계획이다. 본 연구는 음성합성기술의 현재 상태를 분석했지만, 이 기술은 여전히 많은 발전 가능성을 가지고 있다. 향후 연구에서는 음성 합성의 품질 향상과 감정 표현의 개선을 위한 알고리즘 개발에 관한 구체적인 실험을 진행할 것이다. 이를 위해 다양한 음성 라이브러리와 사용자 맞춤형 데이터 셋을 활용하여, 더 정교하고 자연스러운 음성을 생성할 수 있는 방법을 모색할 것이다. 또한, 사용자가 더 직관적으로 창작할 수 있도록, 소프트웨어의 사용자 인터페이스(UI)를 개선하고, 다양한 실험을 통해 실제 음악 창작 환경에서의 유용성을 평가할 계획이다.

둘째, 저작권법 및 법적 규제 개선을 위한 연구가 필요하다. 본 연구에서 언급한 저작권법 관련 이슈는 향후 기술 발전에 중요한 영향을 미칠 것이다. 현재의 저작권법은 음성합성기술에 완벽하게 대응하지 못하며, 이에 따른 법적 분쟁의 가능성도 존재한다. 향후 연구에서는 한국과 중국의 저작권법을 비교하고, 음성합성기술이 포함된 음악 창작물에 대한 법적 해석과 제도적 개선 방안을 구체적으로 제시할 것이다. 또한, 합성 음성이 특정 가수의 음성을 유사하게 모방할 경우 발생할 수 있는 저작권 문제를 해결하기 위해, 합성 음성의 지적 재산권 귀속에 관한 명확한 기준을 마련할 것이다. 이를 통해 법적 보호 체계를 강화하고, 음성합성기술의 발전과 음악 창작이 법적으로 원활하게 조화를 이룰 수 있도록 돕는 정책적 제언을 제시할 예정이다.

셋째, 문화적 다양성과 산업의 다양한 분야에 대한 연구를 확대할 계획이다. 본 연구는 주로 개발자, 사용자, 가창자를 중심으로 인터뷰를 진행했으나, 다양한 문화적 배경과 산업적 관점에서의 음성합성기술에 대한 반응을 다루는 연구가 부족하다. 향후 연구에서는 세계 여러 나라의 음악 창작자 및 소비자들을 대상으로 설문조사 및 인터뷰를 진행하여, 문화적 차이에 따른 음성합성기술의 수용도와 기대에 대해 더욱 풍부한

데이터를 수집할 것이다. 이를 통해 글로벌 음악 산업에서의 음성합성기술의 활용 가능성을 더욱 넓힐 수 있을 것이다.

마지막으로, 정책적 제언과 법적 대응 체계의 구축을 위한 연구도 진행할 예정이다. 음성합성기술의 빠른 발전 속도에 맞춰, 저작권 및 지적 재산권 관련 법률은 끊임없이 변화해야 한다. 특히, 합성 음성의 법적 귀속 문제와 관련하여 기술적 주적 시스템 구축이 중요한 이슈로 떠오르고 있다. 향후 연구에서는 음성합성기술에 대한 법적 대응 체계를 구축하기 위한 기술적 해결책을 제시하고, 대중들에게 저작권에 대한 교육을 강화할 방안을 마련할 것이다. 또한, 관련 법률의 개정을 위한 제언을 통해, 기술과 법적 보호가 함께 발전할 수 있는 정책적 방향을 제시할 것이다.

이와 같은 향후 연구 방향을 통해 음성합성기술의 발전뿐만 아니라, 음악 산업과 저작권법의 법적 체계가 조화를 이루며, 기술과 법이 함께 발전하는 미래를 위한 중요한 기초 자료를 제공할 수 있을 것이다.

참고문헌

- [1] G. M. Arumsari, K. H. Lee, D. Kim, F. M. Lim, S. Almatari, and T. P. Vu, "From Uncanny to Unison: Entertainment Professionals' Perceptions of Singing Voice Synthesis (SVS) and Singing Voice Conversion (SVC) Technologies," in *Proceedings of HCI Korea 2024*, Hongcheon, pp. 785-793, January 2024.
- [2] B. Park, "Application and Technology of Voice Synthesis Engine for Music Production," *Journal of Digital Contents Society*, Vol. 11, No. 2, pp. 235-242, June 2010.
- [3] T. S. Goetze, "AI Art is Theft: Labour, Extraction, and Exploitation: Or, On the Dangers of Stochastic Pollocks," in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, Rio de Janeiro, Brazil, pp. 186-196, June 2024. <https://doi.org/10.1145/3630106.3658898>
- [4] G. Jabour, *Drake Or Fake? Perceptions, Concerns, and Business Implications of AI-Generated Vocals*, Plan II Honors Thesis, The University of Texas at Austin, Austin, TX, April 2024.
- [5] J. L. Flanagan and L. R. Rabiner, *Speech Synthesis*, Stroudsburg, PA: Dowden, Hutchinson and Ross, p. 33, 1973.
- [6] R. Hoffmann, P. Birkholz, F. Gabriel, and R. Jäckel, "From Kratzenstein to the Soviet Vocoder: Some Results of A Historic Research Project in Speech Technology," in *Proceedings of the 20th International Conference on Speech and Computer (SPECOM 2018)*, Leipzig, Germany, pp. 215-225, September 2018. https://doi.org/10.1007/978-3-319-99579-3_23
- [7] M. R. Schroeder, "A Brief History of Synthetic Speech," *Speech Communication*, Vol. 13, No. 1-2, pp. 231-237, October 1993. [https://doi.org/10.1016/0167-6393\(93\)90074-U](https://doi.org/10.1016/0167-6393(93)90074-U)
- [8] J. Q. Stewart, "An Electrical Analogue of the Vocal Organs," *Nature*, Vol. 110, No. 2757, pp. 311-312, September 1922. <https://doi.org/10.1038/110311a0>
- [9] D. H. Klatt, "Review of Text-to-Speech Conversion for English," *The Journal of the Acoustical Society of America*, Vol. 82, No. 3, pp. 737-793, September 1987. <https://doi.org/10.1121/1.395275>
- [10] X. W. Yang, *Song Synthesis Technology and Application Exploration*, Master's Thesis, Nanjing University of the Arts, Nanjing, China, March 2015.
- [11] X. Huang, A. Acero, J. Adcock, H.-W. Hon, J. Goldsmith, J. Liu, and M. Plumpe, "Whistler: A Trainable Text-to-Speech System," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, Philadelphia: PA, pp. 2387-2390, October 1996. <https://doi.org/10.1109/ICSLP.1996.607289>
- [12] Y. E. Kim, *Singing Voice Analysis/Synthesis*, Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, September 2003.
- [13] P. R. Cook, "Singing Voice Synthesis: History, Current Work, and Future Directions," *Computer Music Journal*, Vol. 20, No. 3, pp. 38-46, 1996. <https://doi.org/10.2307/3680822>
- [14] J. O. Smith III and X. Serra, "PARSHL: An Analysis/Synthesis Program for Non-harmonic Sounds Based on a Sinusoidal Representation," in *Proceedings of the 13th International Computer Music Conference (ICMC 1987)*, Champaign-Urbana: IL, pp. 290-297, August 1987.
- [15] J. Liao, "The Musical Function and Aesthetic Representation of Music Created with Vocaloid," *Art Evaluation*, Vol. 13, pp. 7-9, 2020.
- [16] H. Kim, H. Park, J. M. Hong, and J. Kim, "Exploring the Roles of PR Practitioners and Utilization of AI Technology in the AI Era," *Journal of Public Relations*, Vol. 28, No. 1, pp. 207-239, March 2024. <https://doi.org/10.15814/jpr.2024.28.1.207>
- [17] Y. R. Cha, "Artificial Intelligence Strategy for Advertising and Media Industries: Focused on In-depth Interviews," *Journal of the Korea Contents Association*, Vol. 18, No. 9, pp. 102-115, September 2018. <https://doi.org/10.5392/JKC A.2018.18.09.102>
- [18] H.-R. Son and C.-K. Lee, "Artificial Intelligence Utilization Strategies in Design: Insights from Expert Interviews," *Journal of Digital Contents Society*, Vol. 25,

No. 8, pp. 2035-2044, August 2024.
<https://doi.org/10.9728/dcs.2024.25.8.2035>

- [19] Grand View Research. Generative AI in Music Market Size, Share & Trends Analysis Report by Component (Software, Services), by Technology (Transformers, Diffusion Models), by Application, by End Use, by Region, and Segment Forecasts, 2024-2030 [Internet]. Available: <https://www.grandviewresearch.com/industry-analysis/generative-ai-in-music-market-report>.



왕우청(Yujing Wang)

2017년 : 경희대학교 Post Modern 음악학과 (음악학사)

2022년 ~ 현 재: 경희대학교 포스트모던음악학과 석사과정
※ 관심분야 : 음성합성기술(Voice Synthesis Technology), 보컬(Vocal), 싱어송라이터(Singer-songwriter) 등



이철희(Chul-Hee Lee)

2009년 : 경희대학교 Post Modern 음악전공 (음악학사)

2013년 : 경희대학교 아트·퓨전디자인 대학원 (음악학석사)

2020년 : 경희대학교 대학원 (응용예술학박사)

2007년 ~ 2021년: 마인드오픈 음악감독

2021년 ~ 현 재: 경희대학교 포스트모던음악학과 조교수

※ 관심분야 : 컴퓨터음악(MIDI), 작곡(Composition), 사운드디자인(Sound Design) 등