

생성형 인공지능을 이용한 실시간 사운드 시각화

백 상 미¹ · 임 수 연^{2*}

¹경북대학교 디지털미디어아트학과 석사과정

^{2*}경북대학교 예술대학 미술학과 교수

Real-time Sound Visualization using Generative AI

Sang-Mi Baek¹ · Soo-Yeon Lim^{2*}

¹Master's Course, Department of Digital Media Arts, Kyungpook National University, Daegu 41566, Korea

^{2*}Professor, Department of Fine Arts, Kyungpook National University, Daegu 41566, Korea

[요 약]

생성형 인공지능 기술이 등장함으로써 인간의 고유한 영역으로 여겨지던 ‘예술’, ‘창작’이라는 분야가 새로운 국면을 맞이하고 있다. 본 연구는 관람자의 의도가 내재된 음성 정보를 입력받고 생성형 인공지능 기술을 활용하여 실시간으로 영상을 구현하는 시스템을 개발하는 것을 목적으로 한다. 개발한 실시간 사운드 시각화 시스템은 3단계의 프로세스로 이루어진다. 첫 번째 단계에서 입력받은 관람자의 말이나 가사 ‘Sound to Text’ AI를 이용하여 텍스트로 변환하고, 두 번째 단계에서는 생성된 텍스트를 ‘Text to Text’ AI를 통해 이미지 생산을 위한 프롬프트로 확장하고 재생산한다. 마지막으로 ‘Text to Image’ AI를 이용하여 이미지를 자동 생성한다. 본 연구에서 제안한 시스템을 전시장에서 시연한 결과, 인공지능 기술을 이용한 창작이 가능하고 이를 통한 창작자와 관람자 간의 상호 작용이 실현됨을 알 수 있었다. 본 연구는 생성형 AI를 활용하여 창작자와 관람자의 창작 의도를 연결하고 이를 실시간으로 시각화하는 과정을 통하여 기술이 창작의 영역에서 어떻게 활용될 수 있는지 탐구하고, 창작의 본질에 대한 새로운 관점을 제공한다는 의의가 있다.

[Abstract]

With the advent of generative artificial intelligence (AI) technology, areas traditionally considered uniquely human, such as “art” and “creation,” are entering a new phase. This study develops a system that takes voice input embedded with the viewer’s intent and utilizes generative AI technology to visualize images in real-time. The developed real-time sound visualization system consists of three steps. In the first step, the viewer’s spoken words or song lyrics are converted to text using “Sound to Text” AI. In the second, the generated text is expanded and reproduced as a prompt for image production using “Text to Text” AI. Finally, images are automatically generated using “Text to Image” AI. Demonstrating the proposed system in an exhibition revealed that creation using AI is possible, and interaction between the creator and the viewer can be realized through it. This research explores how technology can be utilized in the realm of creation by connecting the creative intentions of creators and viewers using generative AI and visualizing it in real-time, offering a new perspective on the essence of creation.

색인어 : 생성형 인공지능, 시각화, 즉시성, 이미지 생성, 사용자의 의도

Keyword : Generative AI, Visualization, Improvisation, Image Creation, User Intention

<http://dx.doi.org/10.9728/dcs.2023.24.10.2453>



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 31 July 2023; Revised 31 August 2023

Accepted 15 September 2023

*Corresponding Author; Soo-Yeon Lim

Tel: +82-53-950-5684

E-mail: sylim@knu.ac.kr

1. 서론

1-1 연구 배경

4차 산업혁명의 시대를 지나면서 기술은 더욱 고도화되고 그 어느 때보다 급변하는 시대가 도래했다. 2022년 11월 30일 OPEN AI사의 ChatGPT-3.5 모델은 최단 시간 이용자 1억 명을 달성하는 신기록을 세우며[1], 인공지능(AI; Artificial Intelligence), 특히 생성형 AI(Generative AI)에 관한 폭발적인 관심을 불러일으켰다.

생성형 AI는 AI의 분야 중 하나로, 주어진 데이터에서 새로운 데이터를 생성하거나 예측하는 기술을 의미한다. 생성형 AI는 머신러닝(ML; Machine Learning)의 한 부분인 딥러닝(DL; Deep Learning)을 활용한 기술이며 이는 인공지능경망에 기반한다. 생성형 AI는 기존 데이터에 학습된 알고리즘과 모델을 사용하여 이미지, 텍스트, 음악, 비디오와 같은 새로운 콘텐츠를 생성하는 AI 기술이다. 이 모델은 대량의 데이터를 통해 훈련되며, 그 과정에서 특정 패턴과 구조를 학습한다. 이렇게 학습된 패턴을 기반으로 새로운 콘텐츠를 생성할 수 있게 되는 것이다.

예술과 창작은 오랜 시간 동안 인간의 고유한 영역으로 여겨졌다. 예술은 인간의 복잡 미묘한 감정, 다양한 경험, 그리고 무한한 상상력의 표현이었다. 그러나 생성형 AI의 등장으로 이 기존의 패러다임은 크게 흔들리게 되었다. AI는 인간의 창작 영역에 깊숙이 들어와 새로운 예술작품을 만들어내며, 인간만이 가능한 영역이라는 경계를 모호하게 만들었다.

예술은 시대와 문화에 따라 변화해 왔고, 현재는 기술의 발전이 그 변화의 주원인이 되고 있다. 생성형 AI는 예술가의 표현 능력을 확장 시킬 것임은 자명한 일이다. 생성형 AI의 효과적 활용 여부는 예술과 디지털 콘텐츠의 미래를 결정할 것이며, 인간의 예술적 표현 가능성을 더욱 확장 시킬 것이다. 이에 우리는 생성형 AI를 예술 창작의 새로운 도구로 어떻게 활용하고 발전시킬 수 있을지에 대하여 고민할 필요가 있다.

1-2 연구 의의 및 방법

생성형 AI는 기존의 AI 기술들보다 더 쉬운 접근성을 제공한다. 쉬운 접근성이라는 개념은 여러 측면에서의 접근이 용이함을 의미한다. 먼저, 비용적인 측면에서 보면 대부분의 생성형 AI 도구들은 오픈소스로 제공되어 비용의 부담 없이 사용할 수 있다. 또한, 기술적인 장벽이 낮아져 특별한 전문 지식 없이도 사용할 수 있는 사용자 친화적인 인터페이스로 제공된다. 특히, API(Application Programming Interface)의 제공은 이러한 접근성을 더욱 용이하게 한다. 사용자는 API를 사용하여 생성형 AI를 자신의 애플리케이션이나 시스템에 쉽게 통합할 수 있다. API를 이용하면 복잡한 설정이나 추가적인 코드 작성 없이도 AI의 기능을 활용할 수 있게 된다. 더

불어, 유연성 측면에서 다양한 플랫폼이나 환경에서 쉽게 통합하거나 활용할 수 있는 구조로 되어 있으며, 사용자의 요구에 따라 쉽게 확장하거나 수정할 수 있다. 이런 다양한 측면의 쉬운 접근성은 생성형 AI가 기존 AI 기술보다 더 많은 사람이 활용할 수 있도록 만든다.

본 연구에서는 자연스러운 사용자 인터페이스(Natural User Interface)를 위한 실시간 인터랙션을 실현하기 위하여 생성형 AI의 API를 이용한다. 또한 기존의 오디오 데이터 시각화 연구가 오디오 데이터와 그 특성, 예를 들면 주파수, 진폭, 노이즈 등을 시각화하는 데 집중한 것과 달리, 본 연구는 사람의 말(자연어)을 입력받아 그 말의 의미를 해석하고, 이를 영상 이미지로 변환하는 방식을 제안한다. 이러한 접근은 사람의 말에 내재된 의미를 자동으로 시각화하는 새로운 방법을 제시한다는 점에서 본 연구의 중요성과 의의가 있다고 할 수 있다.

이미지를 연상할 수 있는 감성이 담긴 암시적인 텍스트를 접하거나 압축된 이미지와 밀도 있는 표현을 ‘시적이다’라고 한다[2]. 본 연구에서는 사람의 말과 ‘시적’인 노래의 가사를 입력받고, 그 정보를 생성형 AI를 통해 텍스트와 이미지로 해석 및 생성한다. 구현 방법을 정리하면 다음과 같다.

- 1단계: 마이크를 통해 말과 가사를 입력받은 후, ‘Sound to Text’ AI를 활용해 텍스트로 변환한다.
- 2단계: 변환된 텍스트는 ‘Text to Text’ AI를 통해 이미지 생성을 위한 프롬프트로 작성된다.
- 3단계: 작성된 프롬프트는 ‘Text to Image’ AI에 입력되어 이미지를 생성한다.

제안한 시스템의 결과로 생성된 이미지는 ‘Show:보여지다’라는 전시를 통해 관람자들에게 시연되었다.

II. 생성형 AI(Generative AI)

생성형 AI는 사용자가 원하는 대로 결과를 만들어내는 인공지능이다. 기존의 딥러닝 기반 AI 기술이 단순히 이전 데이터를 기반으로 예측하거나 분류하는 정도였다면, 생성형 AI는 이용자가 요구한 질문이나 과제를 해결하기 위해 스스로 데이터를 찾아서 학습하고 이를 토대로 능동적으로 데이터나 콘텐츠 등 결과물을 제시하는 한 단계 더 진화한 AI 기술이다[3]. 생성형 AI 모델은 방대한 양의 데이터에 대하여 모두 습득한 후 해당 데이터의 패턴을 수집하고, 분석하여 새로운 예제를 생성해 낼 수 있다[4]. 생성형 AI는 다양한 분야에 활용되고 있지만, 본 장에서는 예술 및 창작 분야, 특히 ‘Show:보여지다’ 전시에 사용된 생성형 AI에 초점을 맞추어 살펴보고자 한다.

2-1 Text 생성형 AI

생성형 AI 중 가장 잘 알려진 모델이자, 우리가 가장 많이 접한 Open AI사의 ChatGPT는 대표적인 ‘Text to Text’ 생성형 AI이다. GPT(Generative Pre-trained Transformer)는 대량의 자연어 처리 데이터를 바탕으로 트랜스포머(Transformer) 알고리즘을 적용하여 사전 훈련하고 이를 바탕으로 기존 콘텐츠와 유사한 새로운 콘텐츠를 생성하는 인공지능 기술이다[5]. ChatGPT 모델은 대량의 텍스트 데이터에 대해 머신 러닝한 다음 서로 다른 언어로의 변환도 가능하다. 키워드, 주제, 요약 등 텍스트와 관련된 기능을 이용하여 기사를 작성하거나, 시나 소설, 제품 설명서나 계획서 등을 만들어 낼 수 있다[6].

ChatGPT(GPT-3.5)는 2022년 11월 대화형으로 출시되면서부터 대중의 관심을 한 몸에 받았다. 2023년 3월, GPT-3.5 모델을 기반으로 업그레이드된 ChatGPT-4.0은 복잡한 감정의 뉘앙스까지 세밀하게 분석하고, 입력된 자연어에 대하여 창의적으로 응답하거나 해석하는 능력에서 더욱 뛰어나다는 평가를 받는다. 특히, 멀티모달(Multi-Modal) 기능을 통해 여러 모드를 동시 혹은 순차적으로 사용할 수 있다는 점이 ChatGPT-4.0의 특징이다.

본 연구에서는 입력된 말이나 가사의 내용을 분석하고 분석한 내용을 이미지 생성을 위한 프롬프트로 재생산하는데 ChatGPT-4.0을 사용하였다.

2-2 이미지 생성형 AI

이미지 생성형 AI의 발전은 다양한 분야에서 그 효과를 보여주고 있다. 이미지 생성형 AI는 딥러닝과 같은 머신러닝 알고리즘을 사용하여 새로운 이미지를 생성하거나 기존 이미지를 수정하는 기술을 의미한다. 텍스트를 입력받아 사실적이거나 예술적인 디지털 이미지로 생성한다. 이 기술은 2010년대 중반부터 GAN(Generative Adversarial Network)과 같은 알고리즘의 등장과 함께 이미지 인식 및 분류 기술의 발전에 기반하여 빠르게 성장하였다. 2014년에 이안 굿펠로우(Ian Goodfellow)가 제안한 GAN은 이미지 생성 기술의 대표적인 예로, 생성기(generator)와 판별기(discriminator)라는 두 개의 신경망이 서로 경쟁하면서 학습한다[7].

2022년 9월 미국 콜로라도주 주립 박람회 연례 미술 경연 대회(Colorado State Fair Fine Arts Competition)에서 제이슨 앨런(Jason M. Allen)이 이미지 생성형 AI 미드저니(Midjourney)로 만든 작품 ‘스페이스 오페라 극장(Theatre D’opera Spatial)’이 디지털 예술/디지털 이미지 사진 부문에서 1등상을 수상하였다[8]. 이를 통해 AI의 예술적 잠재력이 널리 알려지고, 또 인정받게 되었다. 그림 1은 제이슨 앨런이 1등상을 수상한 작품이다.






그림 1. 스페이스 오페라 극장, 제이슨 앨런
Fig. 1. Theatre D’opera Spatial, Jason M. Allen

이미지 생성형 AI는 텍스트를 입력하면 그에 맞는 이미지를 생성해 주는 ‘Text to Image’ 생성형 AI이다. 최근 이미지 생성형 AI등 중 가장 많은 주목을 받는 모델인 미드저니, DALL-E, 스테이블 디퓨전(Stable Diffusion)의 특징을 비교하면 표 1과 같다.

본 연구는 오픈 소스로 제공되는 스테이블 디퓨전 모델을 활용하여 다양한 이미지 생성 실험을 수행하였다. 연구 과정에서 API 제공 서비스인 ‘컴퓨터렌더(Computerender)’를 이용하여 스테이블 디퓨전 모델의 활용성을 검증하였다.

이미지 생성형 AI는 디지털 아트와 디자인 분야에서 새로운 패러다임을 제시하고 있다. 스테이블 디퓨전과 같은 이미지 생성형 AI 모델의 발전은 새로운 변화를 주도하는 중요한 역할을 하고 있다. 연구자들과 아티스트들은 기술의 발전을 통해 예술의 새로운 가능성을 탐구하게 될 것이고, 발전하는 기술의 활용은 기존의 창작 방식과는 다른 독특하고 혁신적인 작품들이 탄생할 수 있으며, 디지털 콘텐츠의 경계를 넓혀 더 다양하고 풍부한 표현의 변화를 가져올 것이다. 이는 일반 대중에게도 예술과 디지털 콘텐츠에 관한 새로운 시각적 경험과 감동을 제공하게 되어, 예술의 세계를 더욱 풍성하게 만들 것이다.

표 1. 대표적인 이미지 생성형 AI의 특징과 차이점
Table 1. Characteristic features and differences of representative image-generating AI

Type	Characteristic
 Midjourney	<ul style="list-style-type: none"> - Use through Discord - Enter text (Prompt) after ‘/imagine:’ to create an image - Continuous improvement of the image is possible - Combines user’s imagination and creativity to create artwork - Source is not disclosed
 DALL-E	<ul style="list-style-type: none"> - Use of web-based interface - Features for inpainting, allowing addition to existing images, and outpainting to expand created images - Source is not disclosed
 Stable Diffusion	<ul style="list-style-type: none"> - Use of web-based interface - Can be installed on a computer - Source disclosed: Enables creation of a custom image generation program using the Style Diffusion source

III. 생성형 인공지능을 이용한 실시간 사운드 시각화 시스템

본 연구는 화자의 의도를 담은 말이나 가사가 있는 사운드를 입력받아 이를 텍스트로 변환하고, 변환된 텍스트 내용을 표현한 이미지로 시각화하였다. 이러한 ‘Sound to Image’ 시스템을 구축하기 위하여 세 가지의 생성형 AI 모델을 활용하였다.

제안한 시스템을 실시간으로 구현하기 위하여 터치디자이너(TouchDesigner) 소프트웨어를 사용하였다. 터치디자이너는 인터랙티브 비주얼 개발 환경(Interactive Visual Development Environment) 소프트웨어로 노드 기반의 인터페이스를 제공하며 사용자는 다양한 미디어와 데이터 형식 간의 연결을 직관적으로 구성할 수 있다. 또한 사용자의 입력이나 요청에 실시간으로 즉각 반응하며 다양한 시각적 요소를 생성하고 조작할 수 있도록 한다.

실시간 연동이 가능한 터치디자이너의 특성을 이용하여 생성형 AI의 API를 노드로 연결하고, 관람자가 입력하는 사운드를 즉각 텍스트로 변환한 후 실시간으로 관련 이미지를 생성하도록 설계하였다. API 연동은 AI 모델의 복잡한 학습 및 실행 과정을 단순화하며, 클라우드에 기반한 서비스로 제공되는 AI 모델을 활용하여 연구의 실용성과 실시간 사운드의 시각화가 가능하게 했다.

사운드 입력에서 이미지 생성까지의 프로세스를 살펴보면, 관람자가 마이크를 통해 단어나 문장을 말하고 화면에 나타난 텍스트로 확인한 후, 이미지 생성 버튼을 누름으로써 말한 내용이 표현된 이미지가 생성되는 것을 확인할 수 있다. 그림 2는 관람자의 Sound 입력부터 이미지 생성까지의 3단계 과정을 보여준다. 이와 같은 변환 및 생성 과정을 통해 관람자와 생성형 AI 간의 상호작용이 가능한 시스템을 완성하였다.

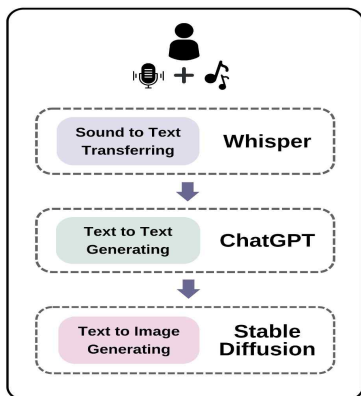


그림 2. ‘Sound to Image’ 진행 과정
 Fig. 2. ‘Sound to Image’ conversion procedure

위 프로세스가 가능하도록 사용된 생성형 AI를 소개하면 다음과 같다.

3-1 Sound to Text

인간은 끊임없이 의사 소통을 하고 무엇인가를 표현하고자 하는 욕구가 있다. 이러한 욕구는 예술 분야의 지속적인 발전을 이끌어왔다. 이 목적을 달성하기 위해 다양한 매체가 사용되었다. 문학은 글자를 통해, 음악은 소리를 통해, 미술은 시각적 요소를 통해, 그리고 무용은 몸짓을 통해 메시지를 전달한다. 매체의 선택은 중요하지만, 그보다 더 중요한 것은 전달하려는 내용의 의도와 메시지이다.

본 연구에서는 ‘마이크를 통해 입력된 음성’과, ‘관람자의 내적 의도를 대신 전달해주는 가사’를 입력하고, 이러한 Sound를 텍스트로 변환하기 위하여 구글에서 발표한 ‘Speech to Text’ 모델인 ‘Whisper’ API를 사용하였다.

‘Whisper’는 OpenAI에서 개발한 자동 음성 인식(ASR: Automatic Speech Recognition) 시스템이다. Whisper는 대량의 음성 데이터를 통해 훈련되었으며, 음성을 텍스트로 변환하는 작업에 사용된다. 다른 생성형 AI는 주어진 입력을 기반으로 새로운 콘텐츠나 정보를 ‘생성’하지만 ‘Whisper’는 입력된 음성을 텍스트로 ‘변환’하여 출력하는 역할을 한다는 점에서 텍스트나 이미지 생성 모델과는 약간 다른 성격을 가진다.

3-2 Text to Text

사운드 입력을 텍스트로 변환한 후, 해당 텍스트에 내재된 의도와 메시지를 정확하게 이미지로 전환하기 위하여 중간 단계를 설정하였다. OpenAI의 ChatGPT는 이 중간 단계에서 핵심적인 역할을 수행하였으며 작가와 화자의 의도를 연결하기 위하여 ChatGPT에 특정 역할을 부여했다. 역할 부여와 함께 변환된 텍스트가 가진 메시지의 본질을 해치지 않으면서 이미지 생성에 적합하도록 그 내용을 묘사하는 프롬프트를 생성하도록 사전에 설정하였다. 이렇게 ChatGPT에 의해 재해석된 프롬프트는 최종 이미지 생성의 기반이 된다.

3-3 Text to Image

전 단계에서 ChatGPT API를 활용하여 관람자로부터의 사운드 입력을 실시간으로 텍스트로 변환하는 작업을 수행하였다. 추가적으로 이 변환된 텍스트는 ChatGPT를 통해 깊게 해석되어 이미지 생성에 적합한 프롬프트로 재구성되었다. 이러한 변환 및 해석 과정을 거친 후, 최종적으로 구성된 프롬프트는 Stable Diffusion 모델에 입력되어 실시간 이미지 생성에 활용되었다. 특히, Stable Diffusion의 ‘Image to Image’ 변환 기능을 활용하여, 작가의 예술적 의도를 반영한 기본 배경 설정과 함께 이미지 생성이 이루어지도록 설계하였다. 이 과정을 통해, 관람자의 표현 의도와 작가의 예술적 비전, 그리고 생성형 AI의 도구적 역할이 통합적으로 작용하는 구조를 구현하였다. 이를 통해 입력된 사운드로부터 이미지까지의 변

환 과정에서 관람자와 작가의 의도가 명확하게 반영되며, 동시에 AI의 기술적 활용이 최적화되는 시스템을 실현하였다.

3-4 Sound to Image 시스템의 의의

본 연구를 통해 구현된 ‘Sound to Image’시스템은 단순히 관람자의 사운드 입력에 기반한 이미지를 생성하는 것만을 넘어, 작가의 예술적 표현 의도 또한 통합되었다는 것이다.

‘Sound to Image’ 시스템은 사운드의 초기 입력부터 최종 출력인 이미지까지의 전 과정에서 두 가지 주요한 차원의 의도가 교차하는 지점이 존재한다. 관람자의 개별적이고 주관적인 표현 의도와 작가의 예술적 표현 의도가 모두 통합되어 있다. 생성형 AI는 이러한 통합을 가능하게 하는 중심적인 도구로 작용한다. 즉각적으로 시각화 되어 생성된 이미지는 단순한 기계적 생성물이 아니라, 인간의 감정과 앞선 기술이 결합되어 생성된 작품이다.

본 연구에서의 주요 도전은, 작가와 관람자, 그리고 서로 다른 특성과 기능을 가진 생성형 AI 모델들 간의 동적인 조화를 가능하게 하는 것이었다. 이러한 상호작용을 통해 연구자의 예술적 표현과 사용자의 내재된 의도가 결합된 독특한 이미지가 형성되었다. 본 연구를 통해 생성형 AI 기술이 실시간으로 사운드를 시각화하는 것을 가능하게 하였으며, 인간과의 상호작용을 통하여 예술과 창작의 도구로서 역할도 충분히 할 수 있다고 판단된다.

IV. 작품 구현 및 전시 분석

한국가곡 ‘연(緣)’의 ‘Show : 보여지다’

4-1 작품 구현

‘생성형 인공지능을 이용한 실시간 사운드 시각화’ 작품을 구현하기 위하여 터치디자이너로 설계된 노드의 구성을 살펴보면 그림 3과 같다. ‘Sound to Text’의 Whisper, ‘Text to Text’의 ChatGPT, ‘Text to Image’의 Stable Diffusion이라는 각각의 생성형 AI의 API가 단계별로 배치되어 연결되면 서 관람자와 작가의 의도가 모두 연결되어 있음을 알 수 있다.

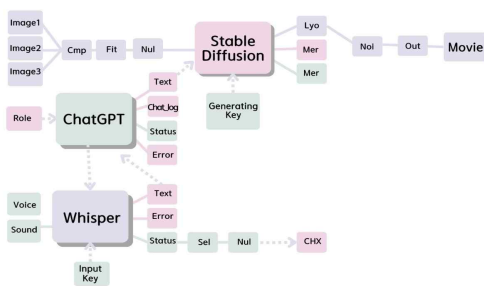


그림 3. ‘Show:보여지다’의 전시 구현을 위한 설계
Fig. 3. Design for ‘Show: To Be Seen’



*한국 가곡의 가사 사용으로 이미지에 한글이 포함되었습니다.
그림 4. 한국 가곡 ‘연(緣)’의 ‘Show:보여지다’의 전시 모습
Fig. 4. Display view of Korean art song ‘Longing (緣), ‘Show: To Be Seen’

한국의 가곡 ‘연(緣)’은 김동현 시인의 시 ‘연’을 가사로 2007년 작곡가 이원주가 작곡하였으며 매우 서정적인 곡으로 깊은 사랑과 이별 후 그리움의 감정을 담고 있다. 가사와 멜로디로 표현된 아련함과 깊은 사랑의 감정, 그리고 그것과 대비되는 그리움과 이별 후의 미련, 쓸쓸함을 동시에 표현되어 있다(그림 4 참조).

이 시의 내포된 의미를 이미지로 보여주기 위하여 연구자는 ‘빛과 그림자’의 이중성을 떠올렸다. 시시각각 변하는 빛에 따라 변화하는 사물과 풍경을 주로 그렸던 인상파 화가 ‘클로드 모네(Claude Monet, 1840~1926)와 같은 위대한 화가(You’re the greatest artist like Claude Monet.)’의 역할(Role)을 ChatGPT에게 부여하였다. 1단계에서 입력된 Text를 3단계 이미지 생성을 위한 프롬프트 생성을 위하여 ‘입력된 메시지를 그림으로 그릴 수 있도록 그 내용을 20단어 이하로 짧지만 자세하게 모습을 상상하여 묘사(Describe the input message in detail, imagining its appearance, so it can be illustrated as a picture, but keep it under 20 words.)’하도록 설정하였다. 생성된 이미지는 사실적이거나, 사진과 같은 느낌을 피하기 위하여 생성된 프롬프트에 ‘Fantastic’과 ‘Dreamy’라는 표현이 함께 작성되도록 설정하였다.

빛나는 사랑의 추억을 표현하기 위해 ‘반짝반짝 빛나는 숲(Twinkling shining forest)’이라는 프롬프트로 생성된 이미지(그림 5. 좌)와 이면에 그림자처럼 숨겨진 아련한 감정과 버리지 못한 미련을 전달하기 위해 대비되는 회색빛의 ‘인상주의적인 붓 획이 아름다운 밤의 꿈결 같은 장면에 생명을 불어넣는다(Impressionistic stroke breathe life in a dreamy and serene scene of a beautiful night)’라는 프롬프트로 생성된 이미지(그림 5. 우)를 Stable Diffusion의 ‘Image to Image’기능으로 생성되는 이미지와 합성되도록 하였다. 이에 더하여 그림 6의 클로드 모네의 작품 중 푸른 하늘과 눈부신 들녘이 담긴 풍경화를 기본 이미지로 제공하였다.

ChatGPT를 사용하여 프롬프트를 재해석하고 Stable Diffusion을 이용하여 이미지를 생성했다. 클로드 모네와 같

은 대표적인 화가의 스타일을 표현기법으로 참조하고, 연구자가 해석한 한국 가곡 ‘연’의 가사에 나타난 빛과 그림자와 같은 양면성을 강조여 최종 이미지가 생성되도록 하였다.



그림 5. ‘Image to Image’에 사용된 배경 이미지 (‘Stable Diffusion’을 통하여 생성)

Fig. 5. Background image used for ‘Image to Image’ (Generated through ‘Stable Diffusion’)

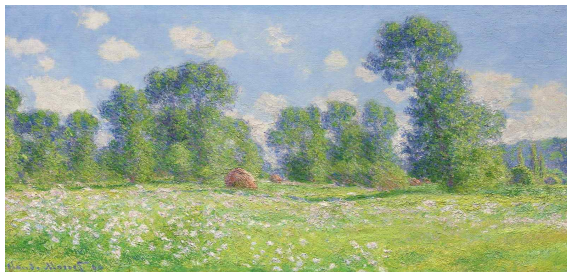


그림 6. 클로드 모네, 풍경화

Fig. 6. Claude Monet, Classic art, Nature, Landscape,

위 과정을 통하여 재해석되고, 생성되어 전시된 한국 가곡 ‘연(緣)’의 가사에 맞는 생성 이미지들은 그림 7과 같다.

전시장에서 관람자는 실시간 사운드의 시각화 인터랙션을 위하여 준비된 마이크를 통하여 직접 이미지로 표현하고 싶은 말(Sound)을 입력하고 이를 바탕으로 생성되는 이미지의 영상화를 직접 체험하도록 하였다. 그림 8은 관람자가 실시간으로 ‘Sound to Image’를 구현한 것으로 관람자 중 1인이 “비 오는 날의 수채화”라는 음성을 입력하였을 때 실시간으로 생성된 이미지(그림8. 좌)와 다른 관람자의 “하늘과 맞닿은 바다”라는 음성을 입력하였을 때 실시간으로 생성된 이미지(그림 8. 우)이다.



*한국 가곡의 가사 사용으로 이미지에 한글이 포함되었습니다.

그림 7. 한국가곡 ‘연(緣)’의 ‘Show:보여지다’의 생성 이미지

Fig. 7. Generated Images of Korean art song ‘Longing (緣)’, ‘Show: To Be Seen’



*한국어 입력으로 이미지에 한글이 포함되었습니다.

그림 8. Voice 입력으로 생성된 이미지

Fig. 8. Generated Image from the Voice input

4-2 전시 분석

본 전시 연구를 통해서 ‘관람자의 표현 의도’와 ‘연구자의 표현 의도’가 생성형 AI를 통해 실시간으로 상호작용하며 시각화하는 과정을 구체화하였다. ‘Show:보여지다’ 전시를 통해 숨겨진 의미나 ‘시적 표현’까지도 인간만이 느끼는 미묘한 감정에 대한 부분에 관한 해석이나, 이미지 생성에 있어서는 아직은 미흡하지만 AI가 실시간으로 구현할 수 있다는 사실을 확인할 수 있었다. 이러한 성과는 기술과 예술 간의 원활한 상호 작용과 협력이 가능함을 시사한다. 아쉬운 점은 현재의 기술 수준에서는 관람자의 Voice 입력 과정이나, Sound의 입력 과정에서 발성 및 발음의 특징에 따라 AI가 입력된 단어나 언어를 정확히 인지하지 못해 입력된 Sound와는 전혀 연관이 없는 결과가 나타나는 경우도 관찰되었다. 이는 사람마다의 독특한 발성 및 발음 특성 때문인 것으로 판단되며, 이를 해결하기 위한 추가적인 기술 연구 및 방법연구가 필요하다는 점을 확인하였다.

V. 결 론

본 연구는 현대 기술의 한계와 그 경계에서의 예술적 표현의 가능성에 대한 탐구로 시작되었다. 그 중심에는 생성형 AI가 있다. 이 연구의 출발점이 된 질문은 ‘생성형 AI가 디지털 콘텐츠의 창작에 어떤 혁신적 변화를 가져올 수 있을까?’, ‘인간만이 가능하다고 여겨지고 있는 예술 및 창작 분야가 과연 AI, 인공지능과의 상호 작용으로 확장하거나 재정의될 수 있을까?’였다.

본 연구는 관람자의 의도가 내재된 음성 정보를 입력받고 인공지능 기술을 활용하여 실시간으로 영상을 구현하는 시스템을 설계하고 구현하였다. 본 연구를 통하여 구현된 실시간 사운드 시각화 시스템은 3단계의 프로세스로 이루어진다. 첫 번째 단계에서 입력받은 관람자의 말이나 노랫말을 ‘Sound to Text’ 생성형 AI를 이용하여 텍스트로 변환하고, 두 번째 단계에서는 생성된 텍스트를 ‘Text to Text’ 생성형 AI를 통해 확장해석하고 재생산한다. 마지막으로 ‘Text to Image’ 생성형 AI를 이용하여 이미지를 자동 생성한다. 구현한 시스템을 전시장에서 시연한 결과, 인공지능 기술을 이용한 창작,

창작자와 관람자 간의 상호 작용이 실시간으로 실현됨을 알 수 있었다.

기술의 발전은 종종 도구의 혁신을 의미한다. 그러나 본 연구를 통해 확인한 바는 기술의 발전이 도구를 넘어 예술과 기술, 그리고 창작자와 관람자 간의 상호 작용이 가능하게 하는 ‘Connection(연결점)’의 역할도 할 수 있다는 사실이다. 생성형 AI를 활용하여 창작자와 관람자의 창작 의도를 연결하고, 이를 실시간으로 시각화하는 과정은 창작의 본질에 대한 새로운 관점을 제시한다. 그 결과, AI의 사용은 단순한 도구의 활용을 넘어서 창작 프로세스의 일부로서 그 가치를 발휘하게 되었다.

또한, 이 연구는 창작의 영역에서 ‘사람만이 가능한 분야’라는 고정관념에 도전하는 계기가 되었다. 생성형 AI의 기술적 발전은 창작자와 관람자의 관계, 그리고 그들이 전달하고 표현하고자 하는 의도의 인터랙션을 통하는 방식으로 예술의 경계를 확장한다. 이러한 변화는 디지털 콘텐츠의 질적 향상은 물론, 창작 과정 자체 혁신을 가능하게 한다. 물론 인공지능 시대의 예술에 대해서는 기대감도 있지만, 생성형 AI를 통해 창작된 창작물에 대한 저작권의 문제나, 생성된 이미지를 도출해 과정에서의 언어적 한계에 관한 문제, 그리고 창작의 영역에 대한 정의 등 함께 고민하고 해결해 나가야 할 문제들도 산재해 있다.

결론적으로, 예술이나 전지콘텐츠, 디지털 콘텐츠 창작 분야에 생성형 AI의 역할은 점점 더 중요해질 것은 자명한 일이다. 예술 및 콘텐츠 창작 분야에서 이 기술의 역할과 확장성은 무한할 것으로 기대하며 본 연구를 통해 그 가능성을 엿볼 수 있었다. 기술의 발전 방향을 잘 파악하고, ‘창작’이라는 영역에 어떻게 적용할 것인지에 대한 고민과 연구를 계속함으로써 기술과 예술이 융합된 디지털 콘텐츠 창작의 더 의미 있는 발전을 기대해 본다. 또한 좀 더 발전된 연구를 통하여 본 연구의 실시간 이미지 생성에 가장 걸림돌이 되었던 ‘발음과 발성’의 차이 및 한계에서 발생하는 Sound ‘인식 오류’에 대한 해결 방법을 모색하고자 한다. 나아가 본 연구를 시작으로 생성형 AI 기술을 다양한 예술 창작 분야에 확장 접목하여 은유적이고 복잡미묘하여 표현이 어려운 내재된 감정을 시각적으로 표현하도록 하여 감정표현에 어려움을 겪는 발달 장애인들의 예술 창작활동이나, 표현에 한계가 있어 의사소통에 어려움을 겪는 사람들에게도 예술성을 표현하도록 도와주는 인공지능 기반의 문화예술교육 플랫폼으로 한걸음 더 나아가 활용될 수 있기를 기대한다.

참고문헌

[1] S. Y. Lee, Opportunity of Generative AI by ChatGPT, KISTI(Korea Institute of Science and Technology Information), Daejeon, S&T Data 2023 Spring Vol. 2, pp. 14-20, April 2023.

[2] J. Y. Kim, “A Study on the ‘Poetic Imagination’ Shown in the Relationship between Text and Image: Focused on ‘Text-Image’, and ‘Image-Text’ in Japanese Advertisement,” *Journal of Basic Design & Art*, Vol. 16, No. 5, pp. 169-186, October 2015.

[3] J. H. Yang and S. H. Yoon, Beyond ChatGPT Transitioning to the Age of Generative AI: Cases of Media and Content Generative AI Services and Strategies for Competitive Advantage, KCA(Korea Communications Agency), Naju, Media Issue & Trend 2023 03+04 Vol. 55 Domestic Report 03, pp. 61-70, April 2023.

[4] StartupN, New Opportunities for Startup “Utilizing the Power of Generative AI” [Internet]. Available: <http://www.newspic.kr/view.html?nid=2023062415011728219&pn=434>.

[5] S. Jang, “ChatGPT has Changed the Future of Writing Education - Focusing on the Response of Writing Education in the Era of Artificial Intelligence,” *Writing Research*, Vol. 56, pp. 7-34, March 2023. <https://doi.org/10.31565/korror.2023..56.001>

[6] Samsung SDS, ChatGPT Technical Analysis White Paper - Part 1: What Is ChatGPT? [Internet]. Available: https://www.samsungsds.com/kr/insights/chatgpt_whitepaper1.html.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, ... and Y. Bengio, “Generative Adversarial Nets,” in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS '14)*, Montréal, Canada, pp. 2672-2680, December 2014.

[8] Seoul Shinmun, Winning an Art Contest with AI-Generated Painting... Plagiarism or Art? [Internet]. Available: <https://www.seoul.co.kr/news/newsView.php?id=20220905019024>.



백상미 (Sang-Mi Baek)

2002년 : 동국대학교 법과대학(학사)

2022년~현 재: 경북대학교 대학원 디지털미디어아트학과 석사

※관심분야 : Digital Art, Interactive Media Installtion, Culture and Arts Education



임수연 (Soo-Yeon Lim)

2004년 : 경북대학교 대학원 컴퓨터공학과(컴퓨터공학 박사)

2018년 : 경북대학교 대학원 디지털미디어아트학과(예술공학 박사)

2014년~2021년: 동양대학교 교수

2021년~현 재: 경북대학교 미술학과 교수

※관심분야 : Digital Art, Interactive Media Installtion, 3D Graphics, Metaverse, HCI (Human Computer Interaction)