

Imaginary Soundscape 기반의 딥러닝을 활용한 회화와 음악의 매칭 및 다중 감각을 이용한 융합적 평가 방법

정하영¹, 김영준², 조준동^{3*}

¹성균관대학교 휴먼ICT융합학과 석사과정, ²성균관대학교 전자전기컴퓨터공학과 박사과정, ³성균관대학교 휴먼ICT융합학과 교수

Convergence evaluation method using multisensory and matching painting and music using deep learning based on imaginary soundscape

Hayoung Jeong¹, Youngjun Kim², Jundong Cho^{3*}

¹Master's Course, Dept. of Human ICT Convergence, Sungkyunkwan University

²Ph. D Course, Dept. of Electrical and Computer Engineering, Sungkyunkwan University

³Professor, Dept. of Human ICT Convergence, Sungkyunkwan University

요약 본 연구에서는 회화 감상에 도움이 되는 사운드시케이프를 구성하기 위해 딥러닝 기술을 활용하여 클래식 음악을 매칭하는 기술을 소개하고 회화와 음악 매칭이 얼마나 잘 되었는지에 대해 평가할 수 있는 평가 지표를 제안한다. 평가 지표는 리커드 5점 척도를 통한 적합도 평가와 멀티모달 측면의 평가로 진행하였다. 회화와 음악 매칭에 대해 13명의 실험 참가자의 적합도 평가의 점수는 3.74/5.0 이었고, 또한 13명의 실험 참가자의 멀티모달 평가에서 회화와 음악 매칭의 코사인 유사도의 평균은 0.79였다. 멀티모달적 평가는 새로운 사용자 경험을 측정할 수 있는 평가 지표가 될 것으로 기대된다. 또한 본 연구를 통해 시각과 청각의 인터랙션을 제한함으로써 다중감각 예술작품 경험을 향상시키고자 하였다. 본 연구에서 제안된 회화와 음악 매칭이 다중감각 예술작품 전시에서 활용되며 더 나아가 이는 시각 장애인들의 예술작품 감상에 대한 접근성을 높일 수 있을 것이라 기대한다.

주제어 : 회화와 음악의 매칭, 딥러닝, 멀티모달 평가, 청각 인터랙션, 다중감각 경험, 사운드시케이프

Abstract In this study, we introduced the technique of matching classical music using deep learning to design soundscape that can help the viewer appreciate painting and proposed an evaluation index to evaluate how well matching painting and music. The evaluation index was conducted with suitability evaluation through the Likert 5-point scale and evaluation in a multimodal aspect. The suitability evaluation score of the 13 test participants for the deep learning based best match between painting and music was 3.74/5.0 and band the average cosine similarity of the multimodal evaluation of 13 participants was 0.79. We expect multimodal evaluation to be an evaluation index that can measure a new user experience. In addition, this study aims to improve the experience of multisensory artworks by proposing the interaction between visual and auditory. The proposed matching of painting and music method can be used in multisensory artwork exhibition and furthermore it will increase the accessibility of visually impaired people to appreciate artworks.

Key Words : Matching of Painting and Music, Deep Learning, Multimodal Evaluation, Auditory Interaction, Multisensory Experience, Soundscape

*This paper was supported by the Korea Research Foundaton Science and Technology Humanities and Social Convergence Research Project in 2018. (2018M3C1B6061353).

*Corresponding Author : Jundong Cho(jdcho@skku.edu)

Received October 6, 2020

Accepted November 20, 2020

Revised October 29, 2020

Published November 28, 2020

1. 서론

최근 국내외에서 시각을 넘어서 청각, 촉각 등의 여러 감각을 통해 예술작품을 감상할 수 있는 다양한 연구 및 전시가 진행되고 있다. 3D 프린팅 제작 기술의 발전으로 예술작품을 부조(relief) 형태로 입체화하여 직접 촉각 작품을 만들 수 있게 함으로써 감상자에게 다중감각 작품 감상 경험을 제공한다[1-4]. 또한 소리, 온도, 진동 패턴이 내장된 벨트를 통해 예술작품을 설계하여 다중감각 작품을 제작한 연구가 있다[5]. 중국 Taizhou Museum의 민속 전시관은 시각, 청각, 후각 등 다중감각 전시를 통해 감상자에게 더욱 인상적이고 현실적인 몰입형 경험을 제공하였으며, 시각, 청각, 후각, 미각 등 여러 감각의 참여는 몰입형 경험을 생성하는 데 중요한 역할을 한다는 연구 결과가 있다[6]. 따라서 본 연구는 인간의 감각 중 청각에 초점을 맞추어 시각 및 청각 인터랙션 기술을 제안함으로써 다중감각 예술작품 경험을 향상시킨다. 사운드스케이프(Soundscape)[7]는 인간이 인식하는 자연적 인위적 음향 환경을 말한다. 본 연구에서는 사운드스케이프 개념을 다중감각 예술작품[1,2]에 적용하여 감상자가 예술작품을 감상할 때 사운드 요소를 통해 새로운 미학적 상상력과 창의력을 자극하는 것을 목표로 한다. 본 논문에서는 예술작품 중 회화에 집중하였고, 회화 감상에 도움이 되는 사운드스케이프를 구성하기 위해 딥러닝 기술을 활용하여 클래식 음악을 선정하는 기술을 소개하고 회화와 음악 매칭이 잘 되었는지에 대해 평가할 수 있는 평가 지표를 제안한다. 이러한 평가 지표는 리커드 5점 척도를 이용한 적합도 평가와 멀티모달 측면의 평가로 구성되었다. 본 연구는 딥러닝을 활용한 청각 인터랙션을 설계하여 시각 및 청각의 다중감각의 융합적 인지를 통해 더욱 풍부한 예술작품 감상 경험을 제공하고자 한다. 나아가 시각 장애인의 다양한 예술작품 전시 관람에 대한 접근성을 높일 것으로 기대된다.

2. 관련 연구

2.1 시각과 청각

감성은 우리의 오감과 밀접한 관련이 있고 감성을 전달하는 감각 기관에서 시각과 청각기관이 대부분을 차지하고 있으며 외부로부터 정보는 시각으로 60%, 청각으로 20%, 나머지 20%는 촉각, 미각, 후각으로 받아들인다[8]. 시각 이미지가 감상자의 감성에 영향을 주어 의미를

전달하고 사운드 또한 감상자의 감성에 영향을 주게 되며, 이때 시각 이미지와 사운드의 감성을 조화시킴으로써 시각 이미지를 극대화할 수 있다[9]. 또한, 시각적 콘텐츠와 짝을 이루는 음악이 감상자의 감정적 경험을 향상시킨다는 연구 결과가 있다[10]. 하지만 이미지와 사운드가 조화롭지 않다면 감상자를 혼란스럽게 만들 것이다. 따라서 본 연구에서는 회화와 어울리는 음악의 매칭을 통해 시각 및 청각 인터랙션의 다중감각 작품 감상을 제안하며 이를 통해 예술작품을 감상할 시 감상자에게 풍부한 감성을 전달하고자 한다.

2.2 다중감각 예술작품

다양한 감각으로 예술작품을 감상할 수 있는 기회가 확대되었다. 네덜란드 빈센트 반 고흐 미술관의 'Feeling Van Gogh'는 예술작품을 3D로 복제하여 촉각 인터랙션과 적절한 향기를 더하여 후각 인터랙션을 제공해 줌으로써 감상자에게 다중감각 예술작품 감상을 경험할 수 있으며 시각 장애인이 예술작품을 감상할 수 있도록 접근성을 높인다. 촉각뿐만 아니라 청각, 후각적 인터랙션이 결합되어 예술작품을 제공할 때 감상자의 몰입감이 향상된다는 연구결과가 있다[11]. 또한 프랑스 레보드프로방스 '빛의 채석장(Carières de Lumières)'과 제주 '빛의 벙커(Bunker de Lumières)[12]'는 미디어 환경 속에서 빛과 음악을 통해 작품을 감상할 수 있는 몰입형 미디어 아트 전시로, 시각을 넘어 예술에 몰입하는 경험을 제공한다. 국내에서는 실로암시각장애인복지관 S 갤러리에서 '다중감각 명화전'을 기획하고 작품의 음성해설, 촉각 패턴 등을 제공하여 시각장애인을 위한 다중감각 예술작품을 전시하였다[5]. 본 연구에서는 시각 및 청각의 인터랙션 감상을 제안하고 딥러닝 네트워크를 통해 회화와 음악을 매칭하는 방법을 소개한다.

2.3 딥러닝 기반의 이미지와 사운드 매칭

예술작품을 감상할 때 음악이 도움이 되거나 방해가 될 수 있기 때문에 사운드스케이프 설계는 매우 중요하다. 따라서 이러한 사운드스케이프를 매칭하는 기술을 선정하는 것도 중요한 문제이다. 최근 딥러닝은 다양한 분야에서 응용되고 있으며 혁신적인 성능 향상을 보이고 있다[13]. 특히 이미지와 사운드를 분석하는 기술을 활용하는 연구가 활발히 진행되고 있다. 딥러닝 기반의 이미지와 사운드의 매칭 방법은 크게 세 가지 방향으로 발전되고 있다. 첫째, 멀티모달의 일관성을 유지하기 위해 이

미지에서 복원 가능한 이미지의 특징점을 추출하여 이를 통해 다시 원본 이미지로 복원 가능한 사운드로 변환하는 방법이다[14]. 둘째, 이미지와 같은 특징점을 추출하기 위해 Semi-supervised Learning 기법을 이용하여 학습시킨 모델을 통해 이미지와 사운드를 매칭하는 방법이다[15,16]. 셋째, 비디오에서 메타데이터를 생성하고 이를 통해 유사한 메타데이터를 가진 이미지와 사운드를 매칭하는 방법이다[17]. 첫 번째 방법은 기존 음악이 아닌 새로운 사운드를 만들어 이미지와 사운드를 매칭하고, 세 번째 방법은 이미지와 사운드에 대한 부가 설명과 같은 메타데이터의 영향을 받아 이미지와 사운드를 매칭한다. 따라서 두 번째 방법을 통해 이미지와 사운드 자체가 가지는 특징만을 이용하여 이미지와 사운드를 매칭시켰다. 하나의 특징적인 오브젝트에 대한 이미지와 사운드는 서로 다른 모달리티를 가지고 있다. 하지만 이를 가리키는 오브젝트는 동일하다. 예를 들어 개의 이미지와 개의 소리는 전혀 다른 데이터의 모달리티를 갖는다. 그러나 실제로는 개라는 하나의 객체를 가리키고 있다. 이 개념을 확장하여 본 연구에서는 이미지와 사운드의 서로 다른 모달리티에 대해 상관성이 잘 표현되는 방향으로 딥러닝을 학습하는 방법인 두 번째 방법을 채택하였다. 본 논문에서는 이러한 딥러닝 기술을 활용하여 회화와 음악의 특징점을 추출하고 회화와 유사한 특징점 분포를 가지는 클래식 음악과 회화를 매칭하였다.

3. 딥러닝 기반의 회화와 음악 매칭 방법

본 연구에서는 사운드 아티스트의 편곡과 작곡을 통해 회화와 어울리는 분위기의 음악을 매칭하는 것이 아니라 딥러닝 네트워크를 통해 이미지와 사운드 자체가 가지는 특징점을 활용하여 회화와 음악의 매칭을 시도하였다.

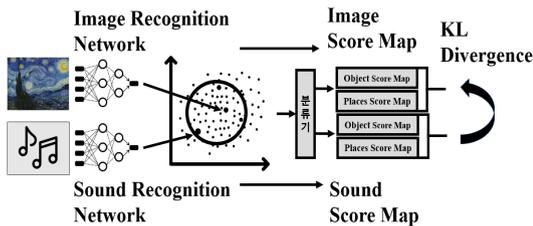


Fig. 1. The System Architecture based on Imaginary Soundscape[15,16]

Fig. 1은 본 연구에서 사용된 딥러닝 네트워크 구조로, Imaginary Soundscape[15] 기반의 딥러닝을 활용하여 회화와 음악을 매칭하는 방법이다. 비디오 데이터에서 이미지 데이터와 사운드 데이터로 분리하고, 이러한 데이터들은 대략적으로 레이블이 지정되고 Semi-Labeled Data로 취급된다. 이미지 데이터는 Image Recognition Network로 입력되는데, 이는 Imagenet과 Places365 데이터 셋으로 사전 학습된 모델이다. 따라서 이미지를 입력받아 이미지의 장면 및 객체, 장소에 대한 특징점을 추출한다. 이와 동시에 Sound Recognition Network는 사운드를 입력받아 장면 및 객체, 장소에 대한 특징점을 추출한다. 이후 KL-Divergence를 이용한 학습 단계에서는 이미지 데이터와 사운드 데이터가 동일한 특징점을 추출할 수 있도록 수행하였다. 이는 딥러닝 네트워크가 이미지와 사운드의 특징점을 추출하여 동일한 벡터 공간에서 표현하도록 학습되었다. 이때 Image Recognition Network는 Mobilenet v3[18]를 사용하였고, Sound Recognition Network는 SoundNet[19]을 사용하였다. 따라서 이 딥러닝 네트워크에 회화를 입력하여 회화의 특징점을 추출하고 이와 동시에 음악 데이터의 특징점을 추출한다. 수집된 음악 데이터베이스 중에 입력된 회화와 가장 유사한 특징점 분포를 가진 음악을 검색할 수 있다. 본 연구에서 사용한 음악 데이터베이스는 1300개의 클래식 음악이다. 이 데이터베이스는 인터넷 사이트를 통해 수집되었다. 본 연구에서 딥러닝 기반의 회화와 음악 매칭을 평가하기 위해 선정한 회화는 빈센트 반 고흐의 '론강의 별이 빛나는 밤에', 빈센트 반 고흐의 '별이 빛나는 밤', 조르주 쇠라의 '그랑드자트섬의 일요일 오후', 르누아르의 '정원에서 파라솔을 든 여인', 빈센트 반 고흐의 '밤의 카페 테라스' 5개의 작품이다. 선정된 회화들은 인상주의 작품으로, 작품 속의 풍경을 음악과 함께 감상함으로써 풍부한 작품 감상 경험을 제공하기에 적합하다고 판단하였다. 회화와 어울리는 음악을 매칭하기 위해 딥러닝 네트워크에 회화를 입력하고 회화의 특징점을 추출하고, 수집한 음악 데이터베이스에서 클래식 음악의 특징점을 추출하여 회화와 유사한 특징점 분포를 가진 상위 3곡을 선정하였다.

4. 평가

3장에서 기술한 회화와 클래식 음악 매칭이 잘 되었는지에 대해 적합도 평가와 미각, 촉각, 온각의 멀티모달 측면의 평가를 진행하였다. 실험 참가자들에게 사전에 상위 3곡의 음악이 딥러닝 네트워크를 통해 선정되었다는 사실을 알리지 않고 실험을 진행하였다. 이는 실험 참가자들이 회화와 음악의 매칭이 잘 되었다는 선입견을 갖지 않고 실험에 참여하도록 하기 위함이다.

4.1 적합도 평가

4.1.1 실험 방법

현재 제주 '빛의 벵커: 반 고흐[12]'에서 전시된 음악과 딥러닝 기반의 회화와 음악 매칭 결과를 바탕으로 적합도 평가를 진행하였다. 적합도 평가는 리커드 5점 척도로 측정하였다. 실험 참가자는 20대의 남자 6명, 여자 7명으로 총 13명이고, 평균 연령은 25.77세(SD=2.36)이다. '빛의 벵커: 반 고흐' 전시에서 빈센트 반 고흐의 '별이 빛나는 밤', '밤의 카페 테라스'에 해당하는 음악을 재생하여 회화와 감상할 수 있도록 하였다. 또한 본 연구에서 사용한 5개의 회화와 딥러닝을 통해 선정된 상위 3곡을 하나씩 재생하여 회화와 음악을 함께 감상할 수 있도록 하였다. 회화와 음악을 감상할 시간을 충분히 준 후 회화와 음악이 얼마나 잘 어울리는지에 대해 리커드 5점 척도로 적합도 평가를 실시하였다.

4.1.2 실험 결과

Table 1은 13명의 실험 참가자를 대상으로 '빛의 벵커: 반 고흐'에 적용된 배경 음악의 적합도 평가 결과이고, Table 2는 딥러닝이 선정한 회화와 음악 매칭 결과의 적합도 평가 결과이다. Table 1, 2에서 "Suitability Evaluation Result"는 실험 참가자들의 적합도 점수의 평균이고, "3 or more percentage"는 5점 척도에서 3점 이상의 점수를 부여한 실험 참가자들의 비율이다.

Table 1의 '빛의 벵커: 반 고흐'에서 '별이 빛나는 밤'에 제공된 배경 음악의 적합도 점수는 3.77점이고 실험 참가자 중 83%가 5점 척도에서 3점 이상의 점수를 주었다. 또한 '밤의 카페 테라스'는 3.85점이고 "3 or more percentage"는 92%로 대부분의 실험 참가자가 3점 이상의 점수를 주었다.

Table 1. Suitability evaluation results of 'Bunker de Lumières'

Painting	Music	Suitability Evaluation Result	3 or more percentage(%)
 Vincent van Gogh 'The Starry Night' 1889, The Museum of Modern Art	Vivaldi - The Four Seasons, Concerto No. 2, Summer	3.77	85
 Vincent van Gogh 'Cafe Terrace at Night' 1888, Kröller-Müller Museum	MILES DAVIS - Ascenseur pour l'échafaud	3.85	85

Table 2. Suitability evaluation results of matching painting and music using deep learning

Painting	Recommended Music	Suitability Evaluation Result	3 or more percentage(%)
 Vincent van Gogh 'Starry Night over the Rhone' 1888, Musée d'Orsay	Peer Gynt - Suite No. 1 Op. 46: III. Anitra's Dance	3.77	100
	Sibelius - Valse Triste, op. 44 no. 1	3.46	77
	Peer Gynt - Suite No. 1 Op. 46: IV. in the Hall of Mountain King	2.1	31
 Vincent van Gogh 'The Starry Night' 1889, The Museum of Modern Art	Beethoven - Piano Sonata No. 3 in C, Op.2	3.92	100
	Prokofiev - 10 Pieces From Romeo And Juliet Op.75 No.9 'Dance Of The Girls With Lilies'	3.69	85
 Georges Seurat 'A Sunday Afternoon on the Island of La Grande Jatte' 1884-1886, The Art Institute of Chicago	Grofe - Grand Canyon Suite, 4. Sunset	3.38	77
	United States Marine Band - Ceremonial Medley	3.38	77
	Tchaikovsky - Swan Lake Op. 20, Act I No. 5, Pas de deux	3.15	85
	Tippett - Concerto for Double String Orchestra 1. Allegro con brio	2.62	38

 <p>Renoir 'Woman with a Parasol in a Garden' 1875, Thyssen-Bornemisza Museum</p>	Sibelius - Valse Triste, op. 44 no. 1	4	100
	Tchaikovsky - Swan Lake, op.20 act 3	2.76	54
	Bach - Brandenburg Concerto No. 6 in B flat major, BWV 1051	2.69	43
 <p>Vincent van Gogh 'Cafe Terrace at Night' 1888, Kröller-Müller Museum</p>	Tchaikovsky - The Sleeping Beauty suite, Op.66a	3.62	92
	Haydn -Symphony No.88 in G Major 1 movement	2.69	54
	Schippers - verdi: nabucco	2.23	38

Table 2에서 전체 회화에서 딥러닝이 선정한 상위 3곡에 대한 적합도 평균은 3.16점이며, 상위 2곡의 평균은 3.45점, 그리고 최상위 1곡(각 회화에 대해 딥러닝이 선정한 회화와 가장 유사한 특징점을 가진 음악)은 3.74점으로 적합도가 높은 것(3.5-4.0)으로 파악되었다. 또한 최상위 1곡에 대한 “3 or more percentage(5점 척도에서 3점 이상의 점수를 준 실험 참가자들의 비율)”는 평균 94%로 높은 비율을 가진다.

빈센트 반 고흐의 ‘별이 빛나는 밤’을 ‘빛의 벙커: 반 고흐’에서 제공된 배경 음악과 딥러닝이 선정한 음악에 대한 사용자 적합도를 비교해 보았다. 딥러닝이 선정한 최상위 적합도를 가진 음악의 적합도 점수는 3.92, ‘빛의 벙커: 반 고흐’에서 제공된 배경음악의 적합도 점수는 3.77 이었다. 한편, ‘빛의 벙커: 반 고흐’에 적용된 ‘밤의 카페 테라스’의 배경 음악과 딥러닝이 선정한 음악의 적합도를 비교해 보았다. 딥러닝이 선정한 최상위 적합도를 가진 음악의 적합도 점수는 3.62, ‘빛의 벙커: 반 고흐’에서 제공된 배경음악의 적합도 점수는 3.85 이었다. 이를 통해 딥러닝이 선정한 음악에 대한 적합도와 ‘빛의 벙커: 반 고흐’에서 작곡가에 의해 작곡된 배경음악에 대한 적합도가 어느 정도 대등한 수준이라는 것을 짐작할 수 있다.

4.2 멀티모달 평가

4.2.1 실험 방법

딥러닝 기반의 회화와 음악 매칭 결과에 대해 멀티모달 측면의 평가를 진행하였다. 실험 참가자는 20대의 남자 4명, 여자 9명 총 13명이고, 평균 연령은 23.69세

(SD=2.58)이다. 이전에 진행한 적합도 평가의 실험 참가자와 중복되지 않도록 하였다. 실험 참가자들은 회화와 음악을 각각 감상하고 난 후 미각, 촉각, 온각별로 연상되는 물체, 상황, 감성 등 자신이 어떻게 느꼈는지를 적도록 하였다. 이후 회화와 음악 매칭이 얼마나 유사한지에 대해 분석하였다. 실험 참가자의 답변을 신순자[19]의 형용사 분류 체계에 따라 분류된 감각 어휘[20]에서 Table 3과 같이 미각, 촉각, 온각의 분류표에 매칭하였다.

Table 3. Sensory Vocabulary

Sensory	Vocabulary
Gustatory Sensory	bitter
	sweet
	salty
	sour
	nutty
	astringent
	spicy
	plain
Tactile Sensory	rough
	smooth
	soft(texture)
	soft(material)
	hard
	moist
Temperature Sensory	sharp
	cold
	cool
	lukewarm
	warm
	hot

회화를 감상하였을 때와 음악을 감상하였을 때의 멀티모달 평가 답변의 유사성을 비교하기 위해 감각 어휘 별 차이하는 답변 비율을 계산하고 코사인 유사도(Cosine Similarity)[21]를 측정하였다. 코사인 유사도는 두 벡터 간 각도의 코사인 값을 이용하여 측정된 벡터 간의 유사도를 나타낸다. 아래 식은 코사인 유사도를 측정하는 수식이다.

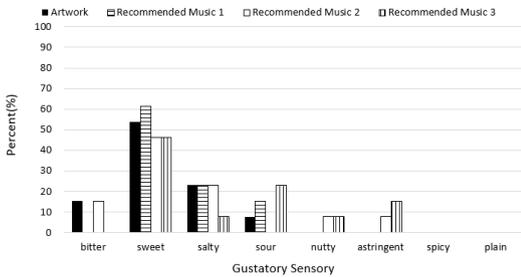
$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

4.2.2 실험 결과

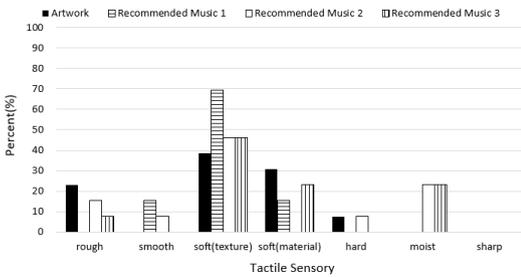
13명의 실험 참가자를 대상으로 회화와 음악 매칭이 잘 되었는지에 대해 멀티모달 측면에서 평가하였다. Fig. 2의 빈센트 반 고흐 ‘별이 빛나는 밤’을 보고 미각, 촉각, 온각에 대해 서술한 답변과 딥러닝이 선정한 상위 3곡을 듣고 미각, 촉각, 온각에 대해 서술한 답변을 비교한 결과는 Fig. 3이다.



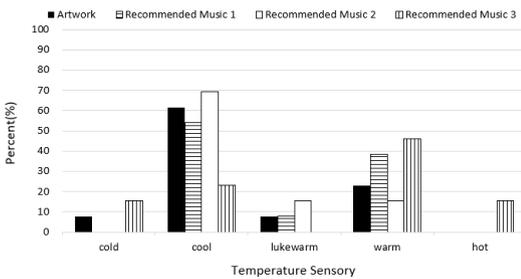
Fig. 2. Vincent van Gogh 'The Starry Night' 1889, The Museum of Modern Art



(a)Gustatory Sensory Evaluation



(b)Tactile Sensory Evaluation



(c)Temperature Sensory Evaluation

Fig. 3. Multimodal Evaluation

빈센트 반 고흐 '별이 빛나는 밤'의 회화를 감상하였을 때와 상위 3곡을 감상하였을 때의 답변을 보면 가장 많은 비율을 차지하는 미각 어휘는 '달다(sweet)', 촉각 어휘는 '부드럽다(soft(texture))'로, 회화와 음악의 최고 비

율의 미각과 촉각 답변이 일치하였다. 하지만, 온각 어휘는 회화 감상과 음악 1과 음악 2에서는 '시원하다(cool)'의 비율이 가장 높았고, 음악 3에서는 '따뜻하다(warm)'의 비율이 가장 많았다. 이는 대체적으로 감각별로 비율이 높은 답변에서 매우 일치하는 것으로 나타났다. 회화 감상에서 미각 어휘의 상위 3개의 답변은 '달다(sweet)', '짜다(salty)', '쓰다(bitter)'였으며, 회화 감상의 미각 어휘의 상위 3개 답변과 음악 감상을 비교하였을 때 음악 1은 67%, 음악 2는 100%, 음악 3은 33%의 일치성을 보였다. 또한, 회화 감상에서 촉각 어휘의 상위 3개 답변은 '부드럽다(soft(texture))', '무르다(soft(material))', '거칠다(rough)'였으며, 음악 감상에서 3곡 모두 회화 감상의 촉각 어휘 상위 3개 답변과 67%의 일치성을 보였다.

또한 회화와 음악 간의 멀티모달 측면의 유사성은 미각, 촉각, 온각별 답변 비율을 코사인 유사도를 활용하여 측정하였다. 이 결과는 Table 3과 같다.

Table 3. Cosine Similarity Measurement Results of matching painting and music using deep learning

Painting	Cosine Similarity
Vincent van Gogh 'Starry Night over the Rhone' 1888, Musée d'Orsay	0.79
Vincent van Gogh 'The Starry Night' 1889, The Museum of Modern Art	0.86
Georges Seurat 'A Sunday Afternoon on the Island of La Grande Jatte' 1884-1886, The Art Institute of Chicago	0.71
Renoir 'Woman with a Parasol in a Garden' 1875, Thyssen-Bornemisza Museum	0.76
Vincent van Gogh 'Cafe Terrace at Night' 1888, Kröller-Müller Museum	0.83

Table 3에서 "Cosine Similarity"는 회화와 상위 3곡간의 미각, 촉각, 온각별 답변의 코사인 유사도 평균이다. 5개의 회화 중 빈센트 반 고흐 '별이 빛나는 밤'에서의 유사도는 0.86으로 상대적으로 가장 높았으며 이는 회화와 음악의 멀티모달 측면에서 가장 유사하다고 할 수 있다. 조르주 쇠라 '그랑드자트섬의 일요일 오후'에서는 0.71로 가장 낮은 유사도를 보인다. 5개의 회화와 전체 상위 3곡에 대해 0.79의 코사인 유사도 평균값을 가지며, 이는 Imaginary Soundscape[15] 기반의 딥러닝이 회화와 음악을 매칭하였을 때, 회화와 음악이 멀티모달 측면에서 보면 어느 정도 유사성을 가진다는 것을 알 수 있다.

5. 결론

본 논문에서는 Imaginary Soundscape[15] 기반의 딥러닝을 활용하여 회화와 잘 어울리는 클래식 음악을 선정하였다. 이후 회화와 음악 매칭이 얼마나 잘 되었는지에 대해 시각 및 청각 감상을 바탕으로 5점 척도의 적합도 평가와 멀티모달 평가를 진행하였다. 멀티모달 평가는 실험 참가자의 답변을 객관적 기준에 매핑하여 해당 비율과 코사인 유사도를 측정하는 방법으로 수행하였으며 이는 새로운 사용자 경험을 제공할 수 있는 평가 지표가 될 것으로 기대된다. 향후 연구에서는 이러한 시각 및 청각 인터랙션이 회화를 감상하고 이해하는 데 얼마나 도움이 되는지 평가하고자 한다. 본 연구의 결과는 시각 장애인을 위한 미술작품 다중감각 감상 시스템인 Blind-Touch[1,2]에 적용하여 회화를 연상시킬 수 있는 BGM(Back Ground Music)으로 사용될 수 있을 것으로 기대한다. 또한, 회화와 음악을 함께 감상할 수 있는 미술관 서비스 애플리케이션에 활용 가능할 것이다.

REFERENCES

- [1] L. Cavazos Quero, J. Iranzo Bartolomé, S. Lee, E. Han, S. Kim & J. Cho. (2018). An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. *In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 346-348). DOI: 10.1145/3234695.3241033
- [2] J. Iranzo Bartolome, L. Cavazos Quero, S. Kim, M. Y. Um & J. Cho. (2019, March). Exploring Art with a Voice Controlled Multimodal Guide for Blind People. *In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 383-390). DOI: 10.1145/3294109.3300994
- [3] J. D. Cho et al. (2019). Color Information Transfer Multi-modal Interface Concept Design for People with Visually Impairment to Appreciate Works of Art - Focused on the Case of "Blind-Touch", a Reproduction Art for Blind -. *Design Works*, 2(2), 44-58.
- [4] Multisensory Artworks exhibition (2019). Human ICT Convergence, (professor: Jun Dong Cho) BlindTouch (Multisensory Painting Platform for the Blind) Exhibition; Exhibition Place: Siloam Center for the Blind S-Gallery.
- [5] D. B. Faustino, S. Gabriele, R. Ibrahim, A. L Theus & A. Girouard. (2017, October). SensArt demo: A multisensory prototype for engaging with visual art. *In Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (pp. 462-465). DOI: 10.1145/3132272.3132290
- [6] S. Wang. (2020). Museum as a Sensory Space: A Discussion of Communication Effect of Multi-Senses in Taizhou Museum. *Sustainability*, 12(7), 3061. DOI: 10.3390/su12073061
- [7] R. Murray Schafer. (1977). *The Soundscape : Our Sonic Environment and the Tuning of the World*. Rochester, Vermont : Destiny Books.
- [8] R. E. Cytowic. (2002). *Synesthesia: A union of the senses*. Cambridge : MIT press.
- [9] Y. G. Jeon. (2004). *(A)Study on the Sound Uses to Maximize Visual Images in Digital Media*. Masters dissertation. Hansung University, Seoul.
- [10] T. Baumgartner M. Esslen & L. Jäncke. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International journal of psychophysiology*, 60(1), 34-43. DOI: 10.1016/j.ijpsycho.2005.04.007
- [11] C. T. Vi, D. Ablart, E. Gatti, C. Velasco & M. Obrist. (2017). Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *International Journal of Human-Computer Studies*, 108, 1-14. DOI: 10.1016/j.ijhcs.2017.06.004
- [12] <https://www.bunkerdelumieres.com>
- [13] K. W. Guk. (2019). Examples of applications by AI technology and industry sectors. *Weekly Technical Trends*, 20, 15-27.
- [14] M. Muller-Eberstein & N. van Noord. (2019). Translating Visual Art into Music. *In Proceedings of the IEEE International Conference on Computer Vision Workshops*. DOI: 10.1109/ICCVW.2019.00378
- [15] Y. Kajihara, S. Ozono & N. Tokui. (2017). Imaginary Soundscape : Cross-Modal Approach to Generate Pseudo Sound Environments. *NIPS Workshop*.
- [16] Y. Aytar, C. Vondrick & A. Torralba. (2016). Soundnet: Learning sound representations from unlabeled video. *In Advances in neural information processing systems* (pp. 892-900).
- [17] A. Sharghi, J. S. Laurel & B. Gong. (2017). Query-focused video summarization: Dataset, evaluation, and a memory network based approach. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4788-4797). DOI: 10.1109/CVPR.2017.229
- [18] A. Howard et al. (2019). Searching for mobilenetv3. *In Proceedings of the IEEE International Conference on Computer Vision* (pp. 1314-1324).
- [19] S. J. Shin. (1999). classification of adjectives. *A*

Collection of Korean Language and Literature Studies at Sookmyung Women's University, 6, 19-40.

- [20] H. W. Jung & K. Nah. (2007). A Study on the Meaning of Sensibility and Vocabulary System for Sensibility Evaluation. *Journal of the Ergonomics Society of Korea, 26(3)*, 17-25.
DOI: 10.5143/JESK.2007.26.3.01
- [21] R. Baeza-Yates & B. Ribeiro-Neto. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.

정 하 영(Hayoung Jeong)

[학생회원]



- 2019년 2월 : 대구가톨릭대학교 의공학과 졸업
- 2019년 3월 ~ 현재 : 성균관대학교 휴먼ICT융합학과 석사과정
- 관심분야 : HCI, 접근성, UX/UI
- E-Mail : lmo3088@skku.edu

김 영 준(Youngjun Kim)

[정회원]



- 2018년 2월 : 한국산업기술대학교 컴퓨터 공학과 졸업
- 2018년 3월 ~ 현재 : 성균관대학교 전기전자컴퓨터 공학과 석박사통합과정
- 관심분야 : 컴퓨터비전, 딥러닝, 객체검출
- E-Mail : yjk931004@skku.edu

조 준 동(Jundong Cho)

[정회원]



- 1993년 6월 : (미) 노스웨스턴대학 박사
- 1995년 3월 ~ 현재 : 성균관대학교 전기전자컴퓨터공학과 교수
- 관심분야 : 체화된 인지, 사용자 경험, 시각장애인 접근성
- E-Mail : jdcho@skku.edu