

머신 러닝 기반 웹소설 맞춤형 사운드 생성 시스템에 관한 연구

이 상 협¹ · 김 주 섭^{2*}

¹서강대학교 아트&테크놀로지학과 석사과정

²서강대학교 아트&테크놀로지학과 교수

Machine Learning-Based Sound Generation System for Web Novels

Sang Hyeob Lee¹ · Jusub Kim^{2*}

¹Master's Course, Department of Art & Technology, Sogang University, Seoul 04107, Korea

²Professor, Department of Art & Technology, Sogang University, Seoul 04107, Korea

[요 약]

전자출판물이 새로운 형태의 독서 경험으로 자리 잡으면서 전자출판물과 시청각 효과를 결합한 다양한 연구가 진행되었다. 관련 연구들은 음악이 독서 경험을 향상시키는 데 도움이 될 수 있다고 보고한다. 그러나 최근 전자출판 시장을 견인하고 있는 웹소설은, 작품마다 음악을 제작하는 데 드는 시간과 추가적인 비용 때문에, 이러한 가능성을 충분히 탐색하지 못하고 있다. 본 논문은 머신러닝으로 생성된 사운드를 활용한, 새로운 웹소설 감상 시스템을 제안하고 그 효과성을 검증하였다. 텍스트 내 각 문장의 감정을 정서 분석 기술로 추측하고, 그 감정에 맞는 배경음악을 머신러닝 기술로 자동 생성하여, 저비용으로 보다 풍부한 독서 경험을 제공하는 것을 목표로 한다. 20대 29명을 대상으로 진행한 실험 결과, 제안한 시스템이 음악이 없는 경우와 비교했을 때, 독서의 몰입도가 통계적으로 유의하게 향상됨을 확인하였다.

[Abstract]

With electronic publication presenting a new form of reading experience, studies have combined electronic publications with audio and visual effects. Related research has shown that music can help enhance the reading experience. However, web novels, the driving force for this market of electronic publication, do not fully explore this possibility, mainly because of the cost to produce music for each web novel. This paper introduces a new reading platform design for web novels using machine-learning-generated sound. The emotion of each sentence in the text is analyzed with sentiment analysis, and background music that matches the emotion is auto-generated to enhance the reading experience. An experiment conducted on 29 participants in their 20s showed statistically significant results in immersion compared to the case without ML-generated music.

색인어 : 웹소설, 다중감각, 사운드 생성, 머신 러닝, UX 디자인

Keyword : Web Novel, Multisensory, Sound Generation, Machine Learning, UX Design

<http://dx.doi.org/10.9728/dcs.2023.24.9.2071>



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 17 February 2023; **Revised** 24 March 2023

Accepted 24 April 2023

***Corresponding Author, Jusub Kim**

Tel: +82-02-705-7976

E-mail: jusub@sogang.ac.kr

I. 서론

웹소설(Web Novel)은 웹에서 출판되어 소비되는 소설을 가리키는 용어로, 주로 압축적이고 흥미 위주의 이야기로 이루어지며, 웹 기반 플랫폼에서 간편하고 저렴한 가격에 구입하여 소비할 수 있는 특징을 지닌다[1].

캐나다의 왓패드(Wattpad), 중국의 치디엔(QiDian), 한국의 네이버 웹소설 등이 대표적인 웹 플랫폼으로, 인터넷 기술과 플랫폼 서비스의 발달로 등장하여, 현재는 스마트폰의 급속한 보급과 함께 전자출판 시장의 주요 콘텐츠로 자리 잡았다. 뛰어난 시각효과가 적용된 스타일과 트렌디한 소재를 바탕으로 엔터테인먼트, 특히 스낵컬처(Snack Culture) 분야에서 빠르게 자리 잡으며 드라마, 영화, 게임 등의 필수 소스가 되고 있다[2]-[4].

웹소설과 전자책 같은 전자출판물의 성장으로, 시각과 사운드 효과가 가미된 전자출판물에 관한 연구들이 진행되어 왔다[5]-[9]. 그중 일부는, 사운드가 독자의 독서 경험을 향상시킬 수 있다는 것을 보여주었다[8],[9]. 그러나 현재 전자출판물 시장에서는 사운드를 제대로 활용하지 않고 있다. 웹소설은 작품의 게시와 감상이 온라인 플랫폼에서 진행되는 특성 때문에, 플랫폼의 환경에 영향을 받는다. 하지만, 현재 대형 웹소설 플랫폼은 사운드 효과를 제공할 수 있는 시스템이 구현되어 있지 않다. 일부 웹소설 작가들은 작품과 어울리는 음악의 링크를 별도로 게시하여, 독자가 작품과 음악을 동시에 감상할 수 있도록 유도했지만, 이러한 방식은 온전한 작품 감상에 번거로움을 발생시킨다. 또한, 이렇게 사용된 음악들도 저작권 관련 문제가 발생할 수 있다. 이로 인해, 작가들은 전문가에게 음악을 의뢰하거나, 작품과 유사한 음악을 직접 찾아 구매해야 한다. 작곡가의 참여하에 자체 사운드트랙이 제작되는 웹소설도 존재하지만, 조회수가 높거나 웹툰 혹은 게임과 같은 2차 콘텐츠로 확장되는 유명 작품 위주로만 진행된다[10]. 이렇듯 작가 개인이 글쓰기에 집중해야 하는 상황에서, 사운드적 요소를 추가하는 작업은 그 자체로 상당한 노력이 필요하고, 추가적인 비용을 발생시키며, 저작권 문제도 고려하여야 한다.

본 연구에서는 저비용으로 웹소설의 읽기 경험을 다중감각적으로 증진시키기 위해, 1) 머신러닝 기술 기반으로 배경 사운드를 자동 생성하는 웹소설 감상 시스템 방식을 제안하고, 2) 프로토타입을 구현하여 그 효과성을 초기 검증하였다. 제안하는 시스템은, 정서 분석 기술로 웹소설 텍스트의 감정을 분석한 후, 머신러닝 기술로 텍스트에 어울리는 사운드를 자동 생성하여, 적절한 지점에서 독자에게 제공한다.

본 연구는 전문 작곡가의 참여 없이 머신러닝 기술 기반의 자동 생성된 사운드를 활용하여도, 독자에게 향상된 읽기 경험을 제공할 수 있다는 것을 실증하였다는 점에 그 의의를 지닌다.

II. 관련연구

2-1 청각을 활용한 전자 출판물 연구

사운드를 활용한 전자 출판물의 효과에 관한 다양한 연구들은 적절한 사운드가 독서 경험을 향상시킬 수 있음을 보여준다. UWE Bristol, Bath Spa University 그리고 University of Birmingham의 연구팀은 '미래의 책'이라는 주제로 Ambient Literature라는 프로젝트를 발표했다[6]. 독자의 모바일 기기 정보까지 활용하는 새로운 전자 출판물 형태로, 프로젝트 결과물 중 하나인 *Breath*는 귀신이 말을 건네는 형태의 작품으로, 독자의 모바일 카메라 및 위치, 시간 정보 등을 시각 및 청각적 효과와 결합하여, 보다 실감 나는 스토리텔링 경험을 전달하려고 하였다[7]. Helsinki School of Economics의 Kallinen은 카페와 같은 야외 환경에서 음악을 들으며 전자기기로 독서할 때, 그 효과에 관한 연구를 진행하였다. 실험 결과, 야외에서 음악을 들으며 독서한 사람이 그렇지 않은 사람보다 높은 만족감과 몰입감을 나타냈다고 검증하였다[8]. Eindhoven University of Technology의 Holenderski와 Hu는 기능적 자기공명영상(fMRI)을 활용해 실제로 음악이 독서에 어떤 영향을 주는지 확인했다[9]. 이를 위해, 피실험자에게 단편 소설을 읽게 하면서, 소설의 내용과 어울리는 음악을 독서 속도에 맞춰 재생하였다. 실험 결과에 따르면, 독자가 사운드트랙을 들으며 전자책을 감상할 때, 음악이 작품의 특정 부분에서 독자의 각성(Arousal) 수준을 높임을 확인하였다. 위와 같은 사례들은, 효과적인 전자 출판물 설계를 위해 청각적인 요소가 중요한 역할을 할 수 있음을 보여준다.

2-2 정서 분석과 시청각적 매핑

정서 분석(Sentiment Analysis)은 텍스트에 표현된 의견과 감정을 컴퓨터로 분석하는 연구를 지칭한다[11]. 모바일 기기의 보급과 SNS(Social Network Service)의 성장으로 사람들의 생각이 온라인에 쌓이면서, 이러한 텍스트 데이터를 다루는 정서 분석이 주목받게 되었으며, 정서 분석 연구는 대량의 텍스트 데이터를 바탕으로 진행되기 때문에 주로 트위터(Twitter)와 같은 SNS 혹은 신문 기사, 블로그 리뷰 기반의 연구가 진행되었다.

정서 분석은 크게 사전 기반 기법(Lexicon based approach)과 머신러닝(Machine-learning approach)으로 나뉜다. 사전 기반 기법은 정서 분석에 사용되는 '정서 사전'을 사용자가 직접 제작하여 적용하는 방법이다. 분석할 텍스트에 '정서 사전'의 단어가 어느 정도 포함되어 있는지 확인한 후, 분석한 값을 합산하여 최종 정서를 분류하는 방법이다. 반면, 머신러닝 기법은 라벨링(Labeling)된 텍스트 정보로 학습을 진행하는 방법이다. 학습이 진행되면서 개별 정서만의

특징을 추출한다. 일반적으로 대량의 문서를 분석할 때는 정서 정보를 학습하여 구축된 정서 분석 모델을 활용하여 진행하기 때문에 머신러닝 기법을 활용한다[12].

정서 분석 결과를 시각화하거나 청각화하여, 정보를 직관적으로 전달하려는 연구들이 진행되었다. *We Feel*은 트위터에 올라오는 대량의 트윗(Tweet)을 정서 분석한 뒤 시각화하여 보여주는 도구이다[13]. 트위터의 위치 정보를 활용하여 개인 혹은 집단의 감정을 실시간으로 분석한 후에 보여줌으로써, WHO(World Health Organization)의 데이터와 같이 연 단위로 발표되는 정보를 활용할 수밖에 없었던 정신 건강 연구자들에게 실시간으로 유용한 정보를 제공할 수 있도록 설계되었다. 국내 유사 사례로는 2017년 진행된 김재민의 연구가 있다[14]. ‘불안’을 기준으로 트위터를 정서 분석한 뒤, 지역별로 불안의 정도를 시각화하여 보여준 연구이다. 사회 문제들을 기존의 연구 방법인 설문조사보다 체계적으로 접근할 수 있도록 설계하였다. 한편, 2011년 발표된 트위터 기반 상태의 청각화 인터페이스 연구[15]는 정서 분석 결과를 시각이 아닌 청각을 활용해서 직관적으로 전달하려는 시도로서 차별점을 지닌다. 인간의 정서를 여섯 가지로 구분한 뒤, 해당 정서와 연관된 단어를 토대로 트윗을 정서 분석하였다. 분석 결과를 사용하여 청각화 작업을 진행하였다. 이를 통해, 사용자가 트위터 내용을 읽지 않고 음의 상승과 하강만을 들으며 타임라인(Timeline)의 함축적인 분위기를 파악할 수 있도록 하였다. 유사 연구로, 2017년 진행된 Seica, M.의 연구가 있다[16]. Plutchik이 정의한 인간의 8가지 기본 감정을 바탕으로[17] 트윗을 정서 분석한 뒤 청각화한 연구이다. 각 감정에 멜로디(Melody), 리듬(Rhythm), 하모니(Harmony)와 같은 음악의 세부 요소들을 할당 및 조절하여 섬세한 표현이 가능하도록 설계하였다.

2-3 머신러닝 기반 사운드 생성

1959년 아서 사무엘(Arther Samuel)이 최초로 정의한 머신러닝은 인공지능의 한 분야로, 명시적으로 프로그램을 작성하지 않고 컴퓨터에 학습할 수 있는 능력을 부여하기 위한 연구를 의미한다[18]. 이는 곧 스스로 데이터에 접근하고 배우는 컴퓨터 프로그램의 개발에 초점을 두고 있다.

머신러닝 기술을 활용한 사운드 연구로는 구글의 마젠타 프로젝트가 대표적이다[19]. 구글이 진행하는 인공지능 예술 프로젝트의 일환으로, 음악과 시각 예술 작품을 창작하는 다양한 프로젝트가 오픈 소스(Open source)로 공개되어 있다. MuseNet[20]은 2019년 미국의 스타트업 OpenAI[21]가 공개한 인공지능 음악 제작 프로젝트이다. 심층 신경망 알고리즘을 사용한 프로젝트로, 4분 정도의 음악을 10개의 악기로 구성하여 제작할 수 있다. *Flow Machines*는 일본 Sony CSL[22]의 프로젝트로, 인공지능 작곡 시스템을 주제로 다양한 서비스를 발표 중이다[23]. 2016년 공개한 노래 *Daddy’s Car*는, 연구소의 머신러닝 알고리즘이 밴드 비틀스

(The Beatles)의 음악을 학습한 후 만들어 낸 음악으로 주목을 받았다. *AI/VA*[24]는 상용화된 머신러닝 음악 서비스이다. 장르와 템포 같은 사용자의 기본 설정을 반영하여 음악이 제작된다. 작곡가 협회에도 정식 등록된 *AI/VA*는 게임과 영화 사운드트랙 같은 다양한 상업 음악 분야에서 활동 중이다.

III. 머신러닝 기반 음악 생성을 활용한 웹소설 감상 시스템 디자인

본 연구에서는, 저비용으로 다중감각을 활용하면서 독자의 읽기 경험 증진이 가능한 웹소설 플랫폼 제안을 위해, 사운드가 독자의 읽기 경험에 효과를 준다는 Kallinen과 Holenderski의 연구를 바탕으로[8],[9], 머신러닝 기술 기반의 사운드 생성 시스템을 디자인하였다. 제안하는 웹소설 감상 시스템은 아래의 다섯 가지 부분으로 이루어진다.

- ① 자연어 처리 알고리즘 기반 웹소설 장면 분리
- ② 장면별로 분리된 텍스트의 감정 추출
- ③ 머신러닝 기반 감정에 어울리는 사운드 생성
- ④ 작품의 장르에 따른 가상악기 설정
- ⑤ 모바일 형식에 기반한 웹소설 인터페이스 디자인

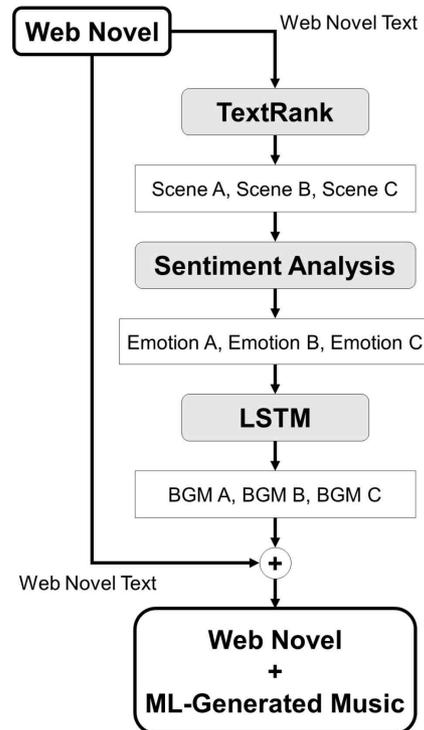


그림 1. 머신러닝 생성 사운드 기반 웹소설 감상 시스템 모식도
 Fig. 1. Web novel reading system with ML-generated sound

웹소설 감상 시스템은 그림 1과 같은 흐름으로 진행된다. 웹소설 작가의 작품은, 자연어 처리 알고리즘을 통해 장면별로 분리되고, 정서 분석 기술을 통해 장면마다 특정 감정으로 분석된다. 분류된 감정을 토대로, 작품과 어울리는 사운드를 머신러닝 기법으로 생성한다. 생성된 배경음악은 다시 작품과 결합한 뒤, 모바일 인터페이스 형태의 웹소설로 독자에게 제공된다.

본 연구의 주요 목적은 머신러닝 기술 기반으로 자동 생성된 사운드로도 독자에게 더욱 몰입감 있는 읽기 경험을 제공할 수 있는지를 확인하는 데 있기에, 새로운 사운드 생성 알고리즘의 개발보다는 현존하는 최신 머신러닝 알고리즘과 데이터셋 활용으로, 배경음악이 자동 생성되어 제공되는 웹소설 읽기 시스템을 구축하고 그 방법의 효과성을 검증하는데 주안점을 두었다.

3-1 자연어 처리 알고리즘 기반 웹소설 장면 분리

머신러닝 기반의 배경 음악 자동 생성을 위해서, 자연어 처리 알고리즘을 활용하여, 다수의 장면으로 웹소설을 분리한다. 이는 영화의 컷(Cut)과 같이 웹소설 한 편을 여러 장면으로 세분화하여, 한 작품 속의 다양한 감정을 세밀하게 연출하기 위함이다.

개발한 시스템 프로토타입에서는, 자연어 처리 오픈 소스 라이브러리인 *Gensim*[25]을 활용하여, 웹소설을 장면들로 분리하였다. *Gensim*은 토픽 모델링, 정보 검색, 유사도 분석 등 머신러닝 알고리즘을 바탕으로 한 자연어 처리 기능을 제공한다. 본 연구에서는, *Gensim*의 *summarize* 기능을 활용하여, 본문의 주요 문장을 추출하였다. 이 기능은 *TextRank* 알고리즘을 통해 요약물 진행하는 방식으로, 문서의 문장과 단어에 가중치를 매겨 요약을 진행하게 된다. *Gensim*의 *summarize ratio* 파라미터값을 조정하여 장면 분류를 위한 주요 문장의 추출 빈도 조절로, 장면의 개수를 조절할 수 있다. 웹소설의 길이에 따라 장면 개수가 증가하도록 설정한다.

3-2 장면별로 분리된 텍스트의 감정 추출

분리된 장면들은 정서 분석 기술을 활용하여 하나의 주요 감정으로 분류된다. 프로토타입에서는 정서 분석을 위해 슬트룩스[26]의 아담(ADAMS) 오픈 AI 서비스 중 "감성&감정 API"를 활용하였다. 해당 API(Application Programming Interface)는 한국어 문장에 대해 정서와 감정 분석 결과를 제공한다. 문장을 입력하면, 기쁨, 신뢰, 공포, 기대, 놀라움, 슬픔, 혐오, 분노 총 8가지 감정 중 밀접한 정서로 분류되고, 0부터 1 사이 수치의 신뢰도와 함께 출력된다.

프로토타입에서는, 각 장면에 포함되는 개별 문장의 정서와 신뢰도 값을 산출하여, 신뢰도 평균이 높은 정서를 해당 장면의 대표 감정으로 선정한다.

3-3 머신러닝 기반 감정에 어울리는 사운드 생성

감정에 영향을 주는 음악적 요소에 관한 연구는 오랫동안 진행되어 왔으며, 음계, 하모니, 리듬, 빠르기, 음고, 멜로디가 주요한 음악적 요소로 연구되어 왔다.

이 중 Meyers, O.[27]는, 음계, 하모니, 리듬, 빠르기 그리고 세기, 총 다섯 가지를 정서에 영향을 주는 음악적 요소로 정의한 후, 사용자 정서 맞춤형 음악 추천 시스템을 제안하였다. 본 연구에서는 이 연구를 기반으로, 해당 다섯 가지 요소를 머신러닝 기반 배경 음악 생성의 파라미터로 활용하여 음악을 생성한다. 각 요소는 감정에 따라 다음과 같은 성질을 지닌다.

- 음계는 긍정적일수록 장조이고(Major), 부정적일수록 단조이다(Minor).
- 하모니는 긍정적일수록 단순하고(Simple), 부정적일수록 복잡해진다(Complexity).
- 리듬은 긍정적일수록 규칙적이며(Regular), 부정적일수록 불규칙적이다(Irregular).
- 빠르기는 긍정적일수록 빠르며(Fast), 부정적일수록 느리다(Slow).
- 세기는 긍정적일수록 높고(High), 부정적일수록 낮다(Low).

위 성질을 토대로, 본 연구에서는, 각 감정에 따른 다섯 가지 파라미터값을 디자인하여 배경 음악을 생성하였다.

프로토타입 개발을 위하여, 구글 마젠타의 머신러닝 작곡 프로젝트 중 하나인 *Performance RNN*[28]을 활용하였다. 해당 프로젝트는 LSTM(Long Short-Term Memory) 알고리즘을 활용한 머신러닝 작곡 프로젝트로, 파라미터 조절을 통해 음악 제작이 가능하다. LSTM은 RNN(Recurrent Neural Network)의 그라디언트 소실 문제(Gradient Vanishing Problem)를 해결하기 위해 제안된 모델로, 이전의 정보를 기억할 수 있는 메모리 셀을 활용하여, 길이가 긴 시퀀스 데이터를 처리할 수 있는 모델이다. 자연어 처리 및 음악 생성과 같은 시계열 데이터 예측 분야에 활용된다. 표 1은 실험을 위해 프로토타입의 작품으로 사용된 웹소설 '전지적 독자 시점'으로부터 추출된 세 가지 주요 감정(신뢰, 혐오, 분노)에 대해, 연구자가 실험을 통해 최종적으로 매핑한 다섯 가지의 음악 파라미터값을 나타낸다. 각 음악 파라미터별 세부 내용은 다음과 같다.

1) 음계(Mode)

음계는 높이의 차례대로 배열한 음의 층계를 의미한다. Schubart와 Steblin은, 기존 클래식 음악을 분석하여 각 조성(Key)과 관련된 인간의 감정을 연구하였다[29],[30]. 이 내용을 바탕으로 감정마다 유사한 조성 세 가지를 선정하였다. *Performance RNN*에서는 *pitch_class_histogram* 변수를 조절하여 생성한다[31].

2) 하모니(Harmony)

화음은 높이가 다른 둘 이상의 음이 동시에 울려서 생기는 합성음을 의미하고, 화성을 뜻하는 하모니는 이러한 화음의 연결을 의미한다. *Performance RNN*의 *notes_per_second* 파라미터는 일정 시간 동안에 들어갈 음의 수를 설정한다. 수치가 커질수록 음이 많아져 화음이 복잡해진다. 긍정적 감정일수록 하모니가 단순하다는 성질에 기반하여 0부터 8까지의 범위 안에서 값을 설정하였다.

3) 리듬(Rhythm)

리듬은 음(음이 없는 상태인 쉬도 포함)이 연속적으로 진행할 때의 어떤 시간적 질서를 의미한다. *Performance RNN*의 *temperature* 변수는 리듬의 복잡성을 표현한다. 변수의 수치가 높아질수록 불규칙한 리듬이, 낮아질수록 안정적인 리듬이 생성되는 성질에 기반하여, 감정에 따라 0.1부터 0.9까지의 범위 안에서 수치를 설정하였다.

4) 빠르기(Tempo)

빠르기는 악곡 진행의 속도를 의미한다. 위에서 언급 Meyers, O.[27]의 연구를 참고하여, 긍정적인 감정은 110~150 Beats Per Minute(BPM)의 범위로 설정하였고, 부정적인 감정은 40~80 BPM의 범위 안에서 값을 설정하였다. 머신러닝 음악 생성 과정의 후반 작업으로 수행하였다.

5) 세기(Loudness)

세기는 소리의 감각적인 크기를 나타내는 척도로서 빠르기와 마찬가지로 음악 생성 과정의 후반 작업으로 수행하였다. *Performance RNN*으로부터 출력된 MIDI 파일의 속도(Velocity) 값 설정에 있어서, 감정이 부정적일수록 음악 세기가 낮다는 성질에 기반하여 긍정적인 감정은 80~100의 범위, 부정적인 감정은 60~80의 범위 안에서 실험을 통해 설정하였다.

수치의 세분화를 위해서 Russell의 감정 차원 모델(Circumplex Model)을 참고하였다. 이 모델은 Arousal(각성)과 Valence(유쾌함/불쾌함)라는 2개의 축을 활용하여 감정을 도식화한 모델이다[32],[33]. 본 연구에서 추출되는 8개의 감정을 Valence 기준으로 긍정과 부정으로 분류한 후, Arousal 상태가 높은 감정일수록 높은 수치의 값을 설정하는 방향으로 진행하였다. 생성 음악의 다양성 확보를 위해, 각 파라미터는 설정 범위 내에서 임의로 선택되도록 설정하였다.

3-4 작품의 장르에 따른 가상악기 설정

본 연구에서는 웹소설의 장르에 따라 차별화되는 가상악기(Virtual Instruments)를 설정하여, 작품의 장르적 표현을 유도한다. 가상 악기란 컴퓨터에서 소프트웨어를 사용하여 하드웨어 악기를 모방하며 연주할 수 있게 한 가상의 악기를 말한다[34]. MIDI(Musical Instrument Digital Interface) 파일의 정보를 바탕으로, 가상 악기 소프트웨어를 활용하여 피

아노 혹은 전자기타와 같은 소리로 출력할 수 있다.

영화 음악이 영상을 보조하듯이, 웹소설의 음악 또한 텍스트를 보조한다. 이에 따라, 가상 악기를 활용하는 영화 음악가, 한스 짐머(Hans Zimmer)의 음악적 특징[35]을 참고하여, 장르별 음악적 표현을 유도할 수 있도록 설정하였다. 표 2는 한스 짐머가 영화의 장르별로 사용한 가상 악기 종류를 바탕으로, 웹소설 장르별로 배치한 가상 악기들을 나타낸다.

표 1. 감정(신뢰, 혐오, 분노)에 따른 음악 파라미터 값 설정 (Performance RNN)

Table 1. Music parameter settings for trust/disgust/anger emotions (Performance RNN)

	Trust	Disgust	Anger
Mode	C Major G Major A Major	D# Minor F Minor G Minor	D# Minor F Minor F# Minor
Harmony	0 - 1	5 - 6	6 - 7
Rhythm	0.1 - 0.2	0.6 - 0.7	0.7 - 0.8
Tempo	110 - 120	50 - 60	60 - 70
Loudness	80 - 85	65 - 70	70 - 75

표 2. 웹소설 장르에 따른 가상악기 설정

Table 2. Virtual instruments settings according to web novel genre

Movie genre	Web novel genre	Virtual instruments
Action, SF	Action, Thriller, SF	String session Electric guitar
Drama, Romance	Romance, Family	String session
History, Adventure	History, Fantasy	String session choir

3-5 장면 전환을 위한 인터랙션 디자인

프로토타입은 웹소설이 주로 소비되는 모바일 기준으로 인터페이스 디자인을 진행하였다.

본 연구에서는, 웹소설의 내용에 따른 머신러닝 기반 자동 음악 생성 시스템의 효과성을 보는 것이 주요 내용이지만, 프로토타입의 주요 특징인, 장면별로 분리된 텍스트와 머신러닝 사운드를 보다 명확히 전달하기 위해, 모바일 기기 사용자의 주요 동작인 밀어 올리기(Swipe-Up) 인터랙션을 장면 전환 방법으로 활용하였다. 밀어 올리는 사용자가 모바일 화면의 아랫부분을 화면 위로 당기는 동작을 말한다. 독자는 밀어 올리기 동작으로 장면과 장면을 이동하고, 종 스크롤(Vertical Scroll) 동작으로 개별 장면을 감상하게 된다. 장면의 전환과 함께, 이전 장면의 배경 음악은 페이드아웃(Fade-Out) 되고, 새로운 장면의 배경 음악이 페이드인(Fade-In) 되도록 설계하였다.

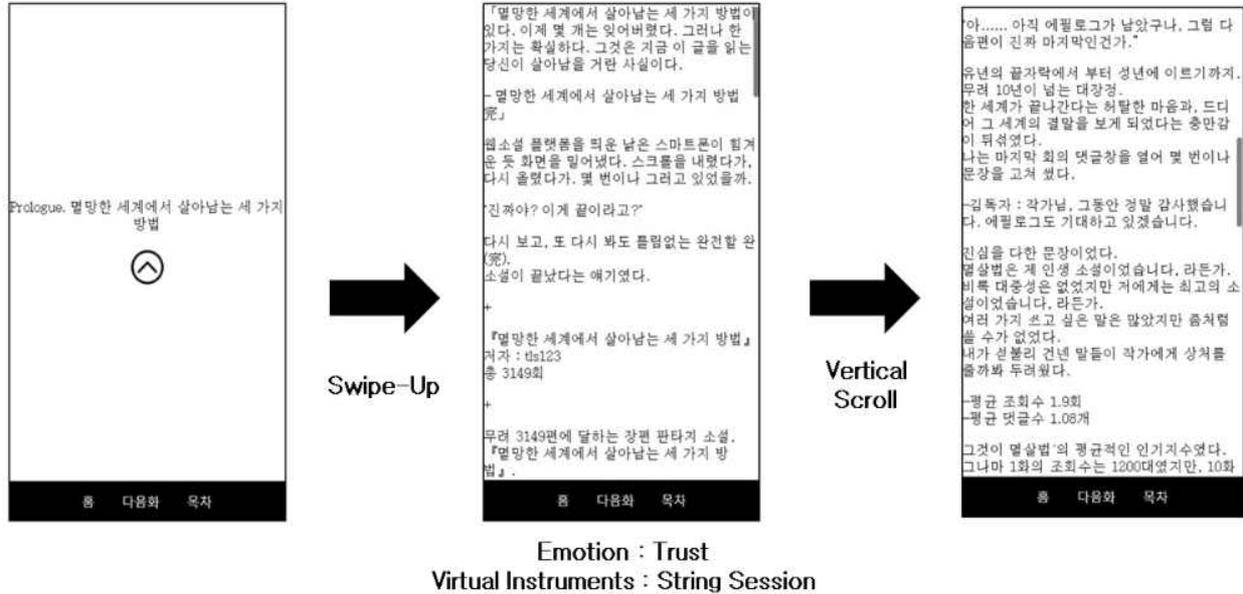


그림 2. 개발된 프로토타입 모바일 화면
 Fig. 2. Mobile screen shots from the prototype

IV. 효과성 검증

4-1 실험 대상 선정 및 실험 설계

기존의 웹소설 플랫폼과 본 연구에서 제안하는 시스템 방식 간에, 독자의 읽기 경험 증진에 있어 어떠한 차이가 발생하는지 그리고 제안하는 시스템의 개별 특징에 관한 평가를 알아보기 위해 실험을 설계하였다.

표 3. 실험 진행을 위한 피실험자 집단

Table 3. Participants' demographic information

	Group A (N=15)		Group B (N=14)	
Age	20-29			
Gender	Male	Female	Male	Female
	6 (40.0%)	9 (60.0%)	6 (42.9%)	8 (57.1%)
Participants who experienced web novel	10(66.7%)		14(100%)	

이를 위해, 웹소설에 친숙한 20대 연령층 29명을 SNS 및 온라인 커뮤니티로부터 모집하여 실험을 진행하였다. 웹소설 사용 경험이 있는 사람들은 24명(82.76%)이었다. 참가자는 같은 작품을 기존 웹소설 플랫폼 형태와 제안하는 프로토타입 형태로 각각 감상한 후 평가를 진행하였다. 감상 순서의 효과를 상쇄하기 위해, 참가자들을 A, B 두 그룹으로 나누어 두 개의 웹소설 감상 순서를 다르게 하여 진행하였다. A그룹은 기존 웹소설 형태의 작품을 감상한 후에 프로토타입을 감

상했고, B그룹은 프로토타입 감상 후에 기존 웹소설 작품을 감상하였다. 한 형태의 작품 감상이 완료될 때마다 개별 평가를 진행하였고, 감상 사이에는 감정 중화 영상을 5분 이상 시청하도록 하여, 다음 작품 감상에 영향이 가지 않도록 하였다.

4-2 실험 평가 항목

읽기 경험에 관한 주제는 지속해 연구되어 왔으나, 이를 검증할 수 있는 합의된 공통의 기준과 문항은 존재하지 않는다. 본 실험에서는 제안하는 시스템의 효과성 검증을 위해 효과성을 다음의 세 가지로 나누어 측정하였다: 몰입성(표 4. A-1, A-2, A-3), 유용성(표 4. B-1), 사용성(표 4. B-2).

표 4. 기존 웹소설 플랫폼과 제안한 프로토타입간 비교 평가를 위한 문항

Table 4. Questions for comparison between established platform and proposed platform

No.	Factors	Questions
A-1	Loss of self-consciousness	I didn't think about anything else while watching web novel.
A-2	Altered sense of time	I didn't know how much time had passed while watching web novel.
A-3	Autotelic experience	It was interesting to appreciate the web novel.
B-1	Useful	This form of web novel has helped to understand the content of the novel.
B-2	Usable	It was easy to read the work in this form of web novel.

몰입(flow)이란 미국 심리학자 칩센트미하이(Mihaly Csikszentmihalyi)가 제시한 개념으로 어떤 행위에 완전히 집중하여 주변에 신경 쓰지 않는 상태이며, 행위 자체가 즐거워서 단지 행위 자체를 목적으로 희생을 감수하려는 상태로 정의된다[36]. 그중에서도 기본적인 광범위한 몰입을 경험할 수 있는 것으로, 독서 활동을 예로 들었다. 본 연구에서는 위의 이론을 토대로, 몰입 수준의 측정 도구를 제한한, 석 임복의 평가 문항을 참고하여 작성하였다[37]. 몰입 측정 관련 문항 중, 자의식의 상실(표 4. A-1 번 문항), 시간 감각의 왜곡(표 4. A-2 번 문항), 자기 목적적 경험(표 4. A-3 번 문항)과 연관한, 세 개 문항을 선정해 작성하였다.

또한, 웹소설 서비스와 연관한 본 연구의 특성상, 서비스 경험에 관한 평가를 위해, 사용자 경험의 해석과 평가를 위한 도구로 UX 연구에서 활용되는, 피터 모빌(Peter Morville)의 허니콤 평가 모델(Honeycomb Model)을 참고하였다. 이 모델은 사용자 경험을 7가지 측면으로 평가할 수 있는 모델로, 본 연구에서는, 유용성(표 4. B-1 번 문항)과 사용성(표 4. B-2 번 문항)으로 두 개의 문항을 구성하였다.

모든 설문 항목은 5점 척도의 리커트 스케일로 구성하였으며, 기존 웹소설과 프로토타입의 개별 감상이 끝날 때마다, 피험자 평가를 받았다. 정성평가를 위해 주관식 답변도 기술하도록 하였다.

또한, 프로토타입 디자인의 개별 요소에 관한 의견을 받기 위하여, 네 가지 추가 문항을 위와 동일한 5점 척도의 리커트 스케일로 구성하였다(표 5). 이 문항들에 대한 설문은 피험자가 프로토타입으로 감상을 완료한 직후에만 실시하였다.

표 5. 프로토타입 디자인의 개별 요소에 관한 평가 항목
Table 5. Questions only for proposed platform features

No.	Factors	Questions
C-1	Scene transitions	I felt that the page method (Swipe-Up) was easy.
C-2	Scene separation	I felt that the number of scenes per episode was adequate.
C-3	ML-generated sound	I felt that background music matches the theme and emotions of web novel.
C-4	ML-generated sound	The background music helped me to immerse myself in watching the story.

4-3 Case Study: '전자적 독자 시점'

제안하는 웹소설 시스템의 효과성 검증을 위해, 유명 웹소설인 '전자적 독자 시점'[38]을 실험 작품으로 선정하였다. 해당 작품은 2018년 1월 6일 웹소설 플랫폼 문피아에서의 첫 연재를 시작으로, 누적 판매 수 1위를 기록하며 2020년 2월 2일 551화로 완결된 소설이다.

약 15분 정도의 독서 시간이 소요되는, 다섯 회차분의 작품을 활용하였다. 해당 작품은 회당 평균 3.4개의 장면으로

분리되어, 다섯 회차 안에서 총 17개의 장면이 추출되었다. 개별 장면의 텍스트는 문장별로 분리된 후, 문장별 감정 분석이 진행되었다. 1화 106문장, 2화 196문장, 3화 147문장, 4화 183문장, 5화 246문장으로 총 878문장이 분석되었고, 장면별 대표 감정이 선정되었다. 최종적으로 신뢰 6장면, 혐오 6장면, 분노 5장면으로 분석되었다.

분석된 결과를 바탕으로 머신러닝 기법으로 사운드를 생성하였고, 제안하는 모바일 인터페이스에 적용하여 프로토타입을 제작하였다. 비교실험을 위해, 같은 작품을 사용한 기존 형태의 웹소설도 제작하였다.

4-4 두 시스템 간 경량 비교 평가 결과

우선, 본 프로토타입의 효과성 검증을 위한 설문에서 사용된 '몰입도' 문항 관련 신뢰도 검증을 위해 척도 신뢰도 검사를 실시하였다. 척도 신뢰도 분석 결과 0.888의 결과로 신뢰도가 있는 것으로 판단하였다(표 6).

표 6. 몰입 척도 신뢰도 검사 결과
Table 6. Scale reliability check results

Number of items	Cronbach's α
A-1, A-2, A-3	.888

본 연구에서 제안한 방법과 기존 웹소설의 사용자 평가에 관한 차이를 보기 위해 대응 표본 t-검증을 진행하였다. 표 7은 5개의 평가 문항에 대한 개별 분석 결과이다.

제안한 프로토타입이 몰입도 3개의 평가 항목에서 모두 통계적으로 유의한 차이를 나타냈다(표 7. A-1 자의식의 상실: $t(28)=-3.84, p=.001$; 표 7. A-2 시간 감각의 왜곡: $t(28)=-2.10, p=.045$; 표 7. A-3 자기 목적적 경험: $t(28)=-3.88, p=.001$). 즉, 본 연구에서 제안한 방법으로 구현한 프로토타입에서 참가자들은 통계적으로 유의하게 높은 몰입감을 느꼈음을 확인하였다.

반면, 참가자들은 내용 이해에 도움을 주었는지 여부에서는 차이를 느끼지 못했고(표 7. B-1 유용성: $t(28)=0.41, p=.682$), 읽는 방식의 편리성 측면에서는 기존의 시스템이 더 편리하다고 보고하였다(표 7. B-2 사용성: $t(28)=4.85, p=.001$)

4-5 프로토타입 디자인 경량 평가 결과

프로토타입 디자인 개별 요소에 관한 5점 척도의 리커트 스케일 평가에서는 장면전환에 사용된 Swipe-Up 방식이 사용자 평가 문항 중 최저점 3.00을 받았다(표 8. C-1 장면 전환). 반면, 배경 음악이 웹소설의 주제나 감정에 부합하였다는 응답은 3.79로 평가 문항 중 상대적으로 최고점을 받았다(표 8. C-3 머신러닝 생성 사운드).

표 7. 기존 웹소설 및 제안한 디자인에 대한 대응 표본 t-검증결과

Table 7. Paired sample t-test result between established platform and proposed platform

Participants (N=29)				
No.	Established	Designed	t	p
A-1	3.59 ± 0.98	4.28 ± 0.59	-3.84	.001**
A-2	3.62 ± 1.24	4.07 ± 0.65	-2.10	.045*
A-3	3.79 ± 0.94	4.41 ± 0.50	-3.88	.001**
B-1	3.93 ± 0.92	3.82 ± 1.00	0.41	.682
B-2	4.07 ± 0.92	3.03 ± 0.73	4.85	.001**

**p<.01, *p<.05

표 8. 프로토타입 특징에 대한 응답 평균

Table 8. Average response to proposed platform features

No.	Questions	Average
C-1	I felt that the page method (Swipe-Up) was easy.	3.00
C-2	I felt that the number of scenes per episode was adequate.	3.69
C-3	I felt that background music matches the theme and emotions of web novel.	3.79
C-4	The background music helped me to immerse myself in watching the story.	3.41

4-6 정성 결과

정성평가를 위해, 실험 참가자에게 프로토타입의 장단점을 물어보았다. 본 프로토타입에서 생성된 배경음악에 대해 긍정적인 의견으로는 독서 경험을 보다 몰입적으로 만들었다는 의견이 가장 많았다.

“처음에는 배경음악이 생소하게 들렸지만 집중하다 보니 상황에 더 잘 어울리는 것 같고 재밌었다.”

“일단 음악 자체는 극 흐름하고 잘 어울렸다. 초반부라 그런가 읽으면서 몰입도 빠르게 되는 데는 음악이 도움을 준 것 같다.”

“음악이 긴장감을 더 키워주고 좋았던 것 같다.”

부정적인 의견으로는 평소 독서 스타일에 따라 배경음악 자체가 오히려 방해 요소로 작용하였다는 의견이 보고되었다.

“글을 조용히 읽는 것을 선호해서 배경음악이 방해 요소로 작용하였다.”

“조용하게 혼자 읽는 공간에서 웹소설을 읽는 것을 즐기는 타입이라 음악이 거슬렸다.”

장면별로 넘어가는 밀어 올리기 방식에 대해서는 부정적인

의견을 얻었다. 기존 웹소설의 종 스크롤 방식과 달라지면서 이질감을 발생시켰으며, 페이지 전환 속도 및 장면의 이동 같은 세밀한 부분에 불편함이 생기면서 발생한 의견으로 추정된다.

“Swipe-Up 실행 시 지연이 있어 불편하다.”

“스вай프 업 방식으로 한 번에 확 넘어가는 형식이 평소 스크롤 스타일과 잘 맞지 않았다.”

그밖에 다크 모드의 부재와 같은 개인별 독서 취향에 관한 세밀한 디자인 요소들에 관한 의견도 보고되었다.

V. 결론

본 연구는, 청각을 활용한 독서가 사용자 경험을 증진한다는 기존의 연구가 존재했음에도, 작품별 음악 제작에 드는 시간과 비용 문제 때문에 서비스 도입을 하지 않는 웹소설 플랫폼의 문제점 인식에서 시작되었다. 이를 위해, 머신러닝 기술을 활용한 웹소설 맞춤형 배경음악 자동 생성 시스템을 제안하였고, 사용자 실험을 통해 그 효과성을 초기 검증하였다. 본 논문에서 제안하는 방법은, 텍스트 분석과 정서 분석 기술을 활용해, 작품을 장면으로 분리한 후, 각 장면의 대표 정서를 분석하여, 분리된 장면의 감정과 어울리는 사운드를 머신러닝 기술로 자동 생성한 뒤, 모바일 인터페이스의 형태로, 해당 텍스트와 함께 독자에게 제공한다.

유명 웹소설 '전지적 독자 시점'을 활용하여, 기존의 웹소설 플랫폼과 제안하는 시스템을 구현한 프로토타입으로, 20대 29명을 대상으로 실험을 진행한 결과, 독서 경험에서의 몰입감이 통계적으로 유의하게 향상됨을 확인하였다. 또한 실험 참가자들의 피드백을 통해, 머신러닝 기술로 생성된 배경음악이지만, 작품의 주제와 감정에 부합한다고 느꼈음을 확인할 수 있었다. 이는 전문 작곡가가 해당 작품을 위해 개별적으로 만든 사운드가 아닌 머신러닝 기반의 자동 사운드 생성 방법으로도, 실제 웹소설 감상에서 다중 감각적 읽기 경험 증진이 가능함을 의미한다.

본 연구는 웹소설 플랫폼에 친숙한 20대 참가자를 대상으로 진행하였기에, 그 결과를 다른 연령대에 일반화시키기는 어려우며, 또한 단일 작품을 대상으로 하였기에, 사례 연구로서의 한계도 지닌다. 또한, 머신러닝 기술로 생성된 음악의 감정이 웹소설 창작자가 본래 의도한 것과 일치하는지 확인하지 못했던 한계점도 존재한다.

하지만 지속해 콘텐츠가 확대되고 있는 웹소설 시장에서, 머신러닝 기술 기반으로 배경음악을 자동 생성한 후, 작품 감상 중 적절한 지점에 음악이 제공되는 방식을 통하여, 저비용으로 사용자 독서 경험을 증진할 수 있다는 것을 실증하였다는 점에, 본 연구의 주요 의의가 있다고 할 수 있다.

향후 연구 주제로는, 텍스트 정서 분석과 배경 음악 자동

생성 간의 매핑에 대한 깊이 있는 연구가 필요하다. 동일한 감정이라도 작품의 전후 흐름 전개를 고려하여 생성하는 방식과, 웹소설 작가가 배경 음악 생성을 정교하게 조정할 수 있는 방식 등에 관한 고려가 필요하다. 아울러 작품에서 자동으로 구분된 장면들 사이의, 화면 및 배경음악의 자연스러운 변환을 위해, 현재의 밀어 올리기 장면 전환 방식을 대체할 수 있는 자연스러운 UI/UX에 관한 연구도 필요하다.

부록

본 연구에서 제안한 시스템의 코드와 감상 시스템을 확인할 수 있는 페이지를 제작하였다. 아래의 URL로 접속하여 확인 가능하다.

URL : <https://github.com/sanghyeob-sogang/prototype>

참고문헌

- [1] KOCCA(Korea Creative Agency). Policy Research for Vitalization of the Web Novel Industry [Internet]. Available: <https://welcon.kocca.kr/ko/info/trend/1945080>.
- [2] The New York Times. Web Fiction, Serialized and Social [Internet]. Available: <https://www.nytimes.com/2014/03/24/technology/web-fiction-serialized-and-social.html>.
- [3] M. J. Ko, "Platform Growth and Possibility of Korean Web Novels," *Korea Digital Publishing Society*, Vol. 13, pp. 32-41, December 2019. <https://doi.org/10.30580/kdips.2019.13.3.32>
- [4] The Korea Herald. Web Fiction Sees Renaissance [Internet]. Available: <http://www.koreaherald.com/view.php?ud=20151022001212>.
- [5] D. J. H. Smeets and A. G. Bus, "The Interactive Animated e-Book as a Word Learning Device for Kindergartners," *Applied Psycholinguistics*, Vol. 36, No. 4, pp. 899-920, July 2015. <https://doi.org/10.1017/S0142716413000556>
- [6] Ambient Literature. This is Your Part of the Story [Internet]. Available: <https://research.ambientlit.com>.
- [7] Ambient Literature. Breathe [Internet]. Available: <https://research.ambientlit.com/index.php/2006/06/30/breathe/>.
- [8] K. Kallinen, "The Effects of Background Music on Using a Pocket Computer in a Cafeteria: Immersion, Emotional Responses, and Social Richness of Medium," in *Proceedings of CHI 2004 Conference on Human Factors in Computing Systems (CHI EA '04)*, Vienna, Austria, pp. 1227-1230, April 2004. <https://doi.org/10.1145/985921.986030>
- [9] K. Holenderski and J. Hu, "Enriching Reading Experience with Dramatic Soundtracks," *Advanced Multimedia Technologies*, Vol. 2, No. 1, pp. 24-31, June 2011.
- [10] Newspim. Kakaopage, OST Release of 'The Legendary Moonlight Sculptor' [Internet]. Available: <https://www.newspim.com/news/view/20200120000210>.
- [11] B. Liu, Sentiment Analysis and Subjectivity, in *Handbook of Natural Language Processing*, 2nd ed. New York, NY: Chapman and Hall/CRC, ch. 26, pp. 627-666, 2010.
- [12] E. J. Myeong, A Sentiment Classification System Using Lexicon-Based Word2vec Features, Master's Thesis, Ewha Womans University, Seoul, August 2018.
- [13] M. E. Larsen, T. W. Boonstra, P. J. Batterham, B. O'Dea, C. Paris, and H. Christensen, "We Feel: Mapping Emotion on Twitter," *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 4, pp. 1246-1252, July 2015. <https://doi.org/10.1109/JBHI.2015.2403839>
- [14] J. M. Kim, Text Corpus Based Emotion Analysis and Visualization Tool, Master's Thesis, Hanyang University, Seoul, February 2018.
- [15] J. Oh, S. Chang, M. Kim, and K. Lee, "An Interface for Sonification of Mood State in Twitter," in *Proceedings of the HCI(Human Computer Interaction) Society of Korea*, Pyeongchang, pp. 140-143, January 2011.
- [16] M. Seiça, R. Lopes, P. Martins, and F. A. Cardoso, "Sonifying Twitter's Emotions through Music," in *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR 2017)*, Matosinhos, Portugal, pp. 586-608, September 2017. https://doi.org/10.1007/978-3-030-01692-0_39
- [17] R. Plutchik, *Emotion, a Psychoevolutionary Synthesis*, New York, NY: Harper & Row, 1979.
- [18] Doopedia. Machine Learning [Internet]. Available: https://www.doopedia.co.kr/doopedia/master/master.do?_method=view&MAS_IDX=160314001525106s.
- [19] Magenta. Google AI [Internet]. Available: <https://magenta.tensorflow.org/>.
- [20] OpenAI. MuseNet [Internet]. Available: <https://openai.com/blog/musenet/>.
- [21] OpenAI. OpenAI [Internet]. Available: <https://openai.com/>.
- [22] Sony Computer Science Laboratories. Sony CSL [Internet]. Available: <https://www.sony CSL.co.jp/>.
- [23] Sony Computer Science Laboratories. Flow Machines [Internet]. Available: <https://www.flow-machines.com/>.
- [24] Aiva Technologies. AIVA [Internet]. Available: <https://www.aiva.ai/>.
- [25] Gensim. Topic Modelling for Humans [Internet]. Available: <https://radimrehurek.com/gensim/>.

- [26] Saltlux. Saltlux [Internet]. Available: <http://www.saltlux.com/index.do>.
- [27] O. C. Meyers, A Mood-Based Music Classification and Exploration System, Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2007.
- [28] Magenta. Performance RNN: Generating Music with Expressive Timing and Dynamics [Internet]. Available: <https://magenta.tensorflow.org/performance-rnn>.
- [29] C. F. D. Schubart, *Ideen zu Einer Ästhetik der Tonkunst*, Hildesheim, Germany, G. Olms, 1969.
- [30] R. Steblin, *A History of Key Characteristics in the Eighteenth and Early Nineteenth Centuries*, Rochester, NY: University of Rochester Press, 1996.
- [31] A. DuBreuil, *Hands-On Music Generation with Magenta*, Birmingham, UK: Packt Publishing, 2020.
- [32] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178, December 1980. <https://doi.org/10.1037/h0077714>
- [33] J. A. Russell and L. F. Barrett, "Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant," *Journal of Personality and Social Psychology*, Vol. 76, No. 5, pp. 805-819, June 1999. <https://doi.org/10.1037/0022-3514.76.5.805>
- [34] PCMag. Virtual Instruments [Internet]. Available: <https://www.pcmag.com/encyclopedia/term/virtual-instrument>.
- [35] Z. Q. Wang, Study on the Function and Form of Hans Zimmer's Film Score in Hollywood Film, Ph.D. Dissertation, Cheongju University, Cheongju, February 2020.
- [36] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, New York, NY: Harper & Row, 1990.
- [37] I. B. Suk and E. C. Kang, "Development and Validation of the Learning Flow Scale," *Journal of Educational Technology*, Vol. 23, No. 1, pp. 121-154, March 2007. <https://doi.org/10.17232/KSET.23.1.121>
- [38] Munpia. Omniscient Reader [Internet]. Available: <https://novel.munpia.com/104753>.

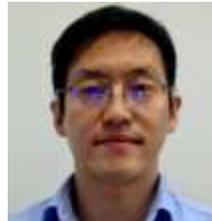


이상협 (Sang Hyeob Lee)

2013년 : 건국대학교
(공학사 - 전자공학)
2021년 : 서강대학교 대학원
(예술공학석사)

2013년 ~ 2018년: LG Electronics

※관심분야 : A.I., 휴먼컴퓨터인터랙션



김주섭 (Jusub Kim)

2000년 : 연세대학교
(공학사-전자공학)
2002년 : 연세대학교 대학원
(공학석사-전자공학)
2008년 : 미국 메릴랜드대 대학원
(공학박사-컴퓨터공학)

2008년 ~ 2012년: 미국 Rhythm & Hues Studios

2012년 ~ 현 재: 서강대학교 아트&테크놀로지학과 교수

※관심분야 : 휴먼컴퓨터인터랙션, 컴퓨터이셔널 미디어