

TV 방송 배경음악 식별/분리/검출용 유사 방송 음악-대사 및 큐시트 데이터셋

김혜미^{1*} · 김성우² · 이상원³ · 김정현⁴ · 박지현⁴ · 유원영⁴

^{1*,4}한국전자통신연구원 통신미디어연구소 콘텐츠연구본부 선임연구원/책임연구원

²인하대학교 전자공학과 학사과정 ³경북대학교 전자공학과 박사과정

Pseudo-Broadcast Music-Speech and Cuesheet Dataset for Background Music Identification/Separation/Detection in TV Broadcast Audio

Hyemi Kim^{1*} · Seongwoo Kim² · Sangwon Lee³ · Junghyun Kim⁴ · Jihyun Park⁴ · Wonyoung Yoo⁴

^{1*,4}Senior/Principal Researcher, Electronics and Telecommunications Research Institute, Daejeon, 34129, Korea

²Undergraduate, Department of Electronic Engineering, Inha University, Incheon, 22212, Korea

³Ph.D Course, School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, 41566, Korea

[요 약]

유사 방송 음악-대사 및 큐시트 데이터셋은 TV 방송에 삽입된 배경음악을 자동으로 식별하는 기술의 성능을 측정하기 위한 데이터셋으로써, 크게 세 가지 요소인 방송 오디오, 배경음악 원본 음원, 그리고 해당 배경음악 원본 음원이 삽입된 시작과 끝 시간 정보를 포함한 큐시트로 구성된다. 이러한 방송 큐시트 데이터셋을 구축하기 위하여 실제 방송 배경음악 편집 과정의 원시 데이터로부터 배경음악 오디오 트랙과 대사 오디오 트랙을 나누어 수집한다. 중복곡을 자동으로 식별하여 제거함으로써 중복 없는 원본 음악 데이터셋을 구축한다. 배경음악 오디오 트랙을 입력으로 음악 식별 기법을 적용하여 큐시트 정보를 얻고, 음악 구간 정보를 추출하여 이를 정제한다. 이 데이터셋은 배경음악을 자동으로 식별하는 기술의 성능을 측정하는 데 사용될 뿐만 아니라 방송 오디오에서 대사와 배경음악을 분리하는 음악-대사 분리 기술과 음악 구간을 검출하는 기술의 성능을 측정하는 데에도 활용할 수 있음을 보인다.

[Abstract]

Pseudo-Broadcast Music-Speech and Cuesheet Dataset is for identifying the background music in TV drama and consists of three parts which are broadcast audios, original background music, and cuesheet including the start and end time at which the original background music is inserted. In order to build it, we collect music and speech track respectively from the raw data generated during the broadcast background music editing process. By automatically identifying and removing duplicate songs, it builds an original music dataset without duplicates. The cuesheet is obtained by applying the music identification technique to the background music audio track as input and refined by extracting music segment information. We show this dataset not only can be used to measure the performance of the background music identification, but also can be used to measure the performance of music-speech source separation and music detection finding music segments in TV broadcast audio.

색인어 : 음악 식별, 음악-대사 분리, 음악 구간 검출, 배경음악, 데이터셋

Keyword : Music identification, Music-speech separation, Music detection, Background music, Dataset

<http://dx.doi.org/10.9728/dcs.2023.24.1.59>



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 07 October 2022; **Revised** 08 November 2022

Accepted 16 November 2022

***Corresponding Author; Hyemi Kim**

Tel: +82-42-860-1816

E-mail: miya0404@etri.re.kr

I. 서론

TV 방송에 삽입된 배경음악을 자동으로 식별하는 기술은 방송사가 지불하는 방송 배경음악 사용료를 투명하고 정확하게 분배하는 데에 핵심이 되는 중요한 기술이다. 이러한 배경음악을 식별하기 위하여 전경음 식별을 목적으로 하는 기존의 오디오 핑거프린트 기술인 랜드마크 기반 오디오 핑거프린팅 기술[1]과 차분 해밍 거리 기반 이진 핑거프린트 기술[2]을 적용할 수 있다. 기존의 오디오 핑거프린트 기술 분야에서는 음악 식별 성능을 측정하기 위하여 먼저 원본 음원 데이터셋을 구축한 후 다양한 형태의 잡음이나 음원의 피치 변환 및 속도 변환 등의 효과를 추가하여 질의함으로써 각각의 변형에 대한 강인성을 측정할 수 있었다. 하지만 TV 방송에 삽입된 배경음악의 식별 성능을 측정할 때는 다양한 변형 효과에 대한 강인성만을 측정하는 것이 아니라 실제 방송 오디오에 포함된 배경음악, 즉, 배우나 출연자의 대사가 추가 되는 상황에서 상대적으로 작은 소리로 혼합된 배경음악을 식별하는 성능을 정확히 측정할 수 있어야 한다.

이를 위해서는 먼저 방송 오디오를 수집하여야 하고 해당 방송 오디오에서 배경음악이 사용된 정보가 정확히 기재된 큐시트 데이터가 필수적이다. 또한 음악 검색 데이터베이스를 구축할 수 있도록 삽입곡의 원본 음악 파일도 필요하다. 이러한 데이터 수집 과정은 여러 장애요인을 내포하고 있다. 현재 배경음악 정산에 사용되는 큐시트는 사람이 직접 청취한 후 해당 음악명을 알아내어 음악이 포함된 시간을 기재하는 방식으로 제작되는데 삽입곡의 정확성 측면에서 한계가 존재한다. 유명한 곡은 비교적 정확히 기재가 가능하겠지만, 연주 중심의 유명하지 않은 음악까지 모두 정확한 곡명을 파악하기는 어려워 삽입된 모든 곡의 곡명을 청취에 의존해 인식하기에는 한계가 있다. 또한 배경음악 가운데는 정식으로 발매되지 않은 곡도 높은 비중을 차지한다. 유명한 곡을 상황에 적합하게 주로 사용하는 예능 분야와는 달리, 드라마 분야에서는 해당 콘텐츠에 어울리는 배경음악을 직접 제작하고 이중 상업성이 높은 일부 음악만을 발매하는 것이 일반적이다. 따라서 방송 오디오에 포함된 모든 원본 음원을 빠짐없이 수집하여 검색 데이터베이스를 구축하기 어렵다.

우리는 기존의 큐시트 데이터 수집에 의존한 방식에서 벗어나 방송 배경음악 식별 성능을 정확히 측정하기 위하여 유사 방송 음악-대사 및 큐시트 데이터셋을 구축하고자 한다. 방송 프로그램은 다양한 장르를 가지는데 앞서 서술한 것처럼 예능의 경우 대부분 기존에 발매되거나 유명한 곡을 삽입하는 데 반해, 드라마의 경우 드라마에 어울리는 분위기의 배경음악이나 배우들의 테마 음악, OST 등을 직접 작곡하여 드라마에 삽입하는 것이 일반적이다. 본 논문에서는 방송 프로그램 장르 중 드라마 배경음악 큐시트 데이터셋을 구축하기 위하여 드라마 배경음악을 직접 작곡하고 편집한 음악 프로덕션으로부터 방송 배경음악을 삽입하는 일련의 작업 과정에서 축적된 원시 데이터를 수집하였다. 이 데이터는 방송가에서

흔히 사용되는 전문가용 비디오 편집 프로그램인 베가스[3]를 기반으로 축적되었다. 베가스 프로젝트 파일에는 여러 오디오 트랙이 포함되어 있는데, 배우들이 연기하는 현장에서 직접 녹음한 대사 위주의 현장 녹음본 트랙과 배경음악으로 삽입된 각 음원들에 다양한 효과가 적용된 다수의 개별 음원 트랙, 그리고 매 시점의 음원 트랙이 모두 혼합된 배경음악 트랙이 포함되어 있다. 오디오 트랙과 함께 배경음악 큐시트 정보를 제공하는 데 필수적인 원곡 음원도 사용된 음원이 중복 및 빠짐없이 포함되도록 사전에 정제하여 데이터셋으로 구축한다. 마지막으로 해당 오디오 및 원곡 음원과 동기화된 배경음악 큐시트 데이터를 구축한다. 배경음악 큐시트는 오디오에서 음악의 시작 시각과 끝 시각 정보가 동기화되고, 원곡 음원과 음악 아이디 정보가 동기화되어야 한다. 데이터셋 구축에 활용된 베가스 프로젝트 예시를 그림 1에 나타내었다. 빨간 사각형 내 여러 파일명이 보이는 부분이 원본 음원에 관한 영역이고, 주황색 영역의 긴 오디오 트랙이 대사가 포함된 현장 녹음본 트랙 및 배경음악 트랙이다. 주황색 영역의 아래에는 최종 배경음악 트랙을 만들어내기 전에 원곡 음원마다 원하는 시점에 삽입한 개별 트랙이 배치되어 있다.

본 논문의 2장에서는 방송 오디오에서 음악과 음성 데이터셋에 관련한 기존 연구를 서술한다. 3장에서는 유사 방송 음악-대사 및 큐시트 데이터셋을 구축하는 방법 및 구축된 데이터셋에 대해 서술한다. 4장에서는 구축된 데이터셋을 방송 배경음악 모니터링에 관련된 세 가지 기술 분야에 적용한다. 먼저 큐시트 정보를 활용하여 배경음악 식별 성능을 측정한다. 다음으로 대사 트랙과 배경음악 트랙을 활용하여 음악-대사 신호 분리 기술의 성능을 측정해보고, 음악이 아닌 대사 오디오 신호를 제거한 후 배경음악 식별 성능을 측정하여 신호 분리 기법이 성능 개선에 미치는 영향을 살펴본다. 마지막으로 큐시트에 포함된 음악의 시작 및 끝 시각 정보를 활용하여 음악이 포함된 구간을 찾는 음악 구간 검출 기법의 성능을 측정하는 실험에 관해 서술한다.

II. 관련 연구

최근 들어 TV 또는 라디오 방송 오디오의 음악과 음성에 관련된 데이터셋이 연이어 발표되고 있다. 먼저 TV 방송에 관한 데이터셋 가운데 TVSM 데이터셋[4]은 TV 방송 프로그램 오디오에서 음성과 음악 구간 정보를 표기한 데이터셋으로 최근에 공개되었다. 데이터셋은 총 3가지 형태로 구성되어 있는데, 먼저 TVSM-cuesheet와 TVSM-Psuedo는 각각 큐시트와 사전 학습된 모델에 기반한 잡음 포함(noisy label) 데이터셋으로 각각 약 55시간과 1,540시간의 대규모 데이터셋이다. TVSM-test는 사람이 직접 오디오를 들으며 음악과 음성 구간을 표기한 상대적으로 정확도가 매우 높은 데이터셋으로 약 15시간 규모이다.

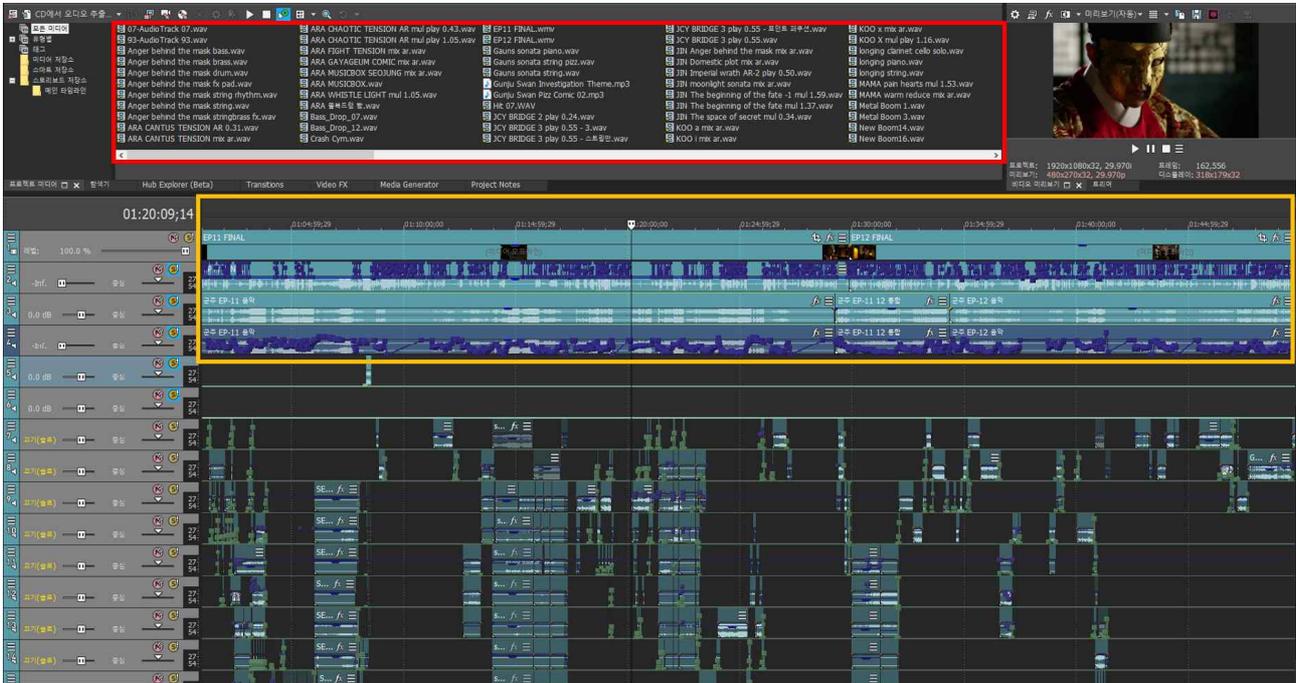


그림 1. 베가스 프로젝트 구성 예시 (원본 음원(빨간색 영역), 대사 및 음악 트랙(주황색 영역))

Fig. 1. An example of a Vegas project configuration (original source files(red box), speech and music tracks(orange box))

데이터셋은 저작권 문제로 오디오를 직접 공개할 수 없어 특정 정보만을 제공하는데, 멜 주파수를 기반으로 주파수와 시간축 정보로 나타낸 멜 스펙트로그램과 VGG 모델로부터 얻은 VGGish 특징, 그리고 MFCC(Mel Frequency Cepstrum Coefficient) 파일로 제공한다. 이 데이터셋은 음악 구간과 음성 구간 예측을 목적으로 하는 최초의 오픈소스 기반 대규모 데이터셋으로 음악이나 음성 구간을 예측하는 데에는 활용 가치가 매우 높으나, 방송 배경음악 식별을 위한 원곡 음원의 메타 정보를 포함하지 않아 배경음악 식별 성능을 측정하는 데에는 활용할 수 없다.

라디오 팟캐스트에 관한 데이터셋 가운데 PodcastMix 데이터셋[5]은 음악과 음성을 혼합하여 구축한 데이터셋으로 음악과 음성 트랙을 각각 제공함으로써 팟캐스트 오디오에서 음악과 음성 오디오를 각각 분리하는 연구에 활용할 수 있다. 기존의 음성 데이터셋인 VCTK dataset[6]과 기존의 음악 데이터셋인 Jamendo CC music[7]을 자체적으로 혼합하여 약 41 시간 규모의 합성 데이터셋을 구축하였는데 전문적으로 제작한 라디오 팟캐스트와는 품질 면에서 차이가 있을 수 있고, 역시 TVSM 데이터셋[4]과 마찬가지로 삽입된 배경음악의 정보가 포함된 큐시트를 제공하지 않는다.

배경음악 식별을 위한 데이터셋으로는 최근 공개된 BAF 데이터셋[8]이 있다. 로열티 분배에서 필수적인 방송 배경음악 식별에 관한 최초의 데이터셋이다. 2,000곡이 배경음악으로 사용되었고 질의 오디오는 23개국, 203개 채널의 실제 TV 방송에서 발췌한 약 57시간 분량의 1분 길이 3,425개 오디오이다. 8kHz로 샘플링되어 고음질 오디오는 아니지만, 오

디오 특징 정보가 아닌 오디오 파일 자체를 제공한다는 점에서 사용성이 높은 데이터셋이라고 할 수 있다. 각 질의 오디오에서 삽입된 배경음악의 시작과 끝 시점을 포함한다. 6명이 각 질의 오디오에 태깅을 하고 한 질의 당 세 명이 교차 태깅을 하여 태깅 신뢰성을 높였으나, 동일 질의에 대해 표기자에 따라 중복되거나, 서로 다른 음악 구간 정보를 표기한 것을 취합한 형태로 제공하고 있어, 활용 시 추가 정제가 필요할 수 있다. 실제 방송에서 발췌한 1분 길이의 질의 오디오와 오디오에 포함된 원본 음원, 그리고 이들에 동기화된 큐시트를 제공함으로써 방송 배경음악 식별 성능 측정이 가능한 최초의 공개 데이터셋이다.

본 논문에서는 BAF 데이터셋[8]과 같이 방송 배경음악 식별이 가능한 유사 방송 음악-대사 및 큐시트 데이터셋을 구축한다. 먼저 현장에서 녹음한 대사 중심의 현장 녹음본 트랙과 음악 편집 프로그램에서 삽입된 배경음악만을 포함하는 배경음악 트랙을 각각 모노 오디오로 렌더링하고 두 트랙을 혼합한 총 세 종류의 오디오 트랙을 수집한다. 다음으로 삽입된 원본 음원 데이터셋을 구축한다. 마지막으로 해당 오디오 트랙에 동기화된 배경음악 큐시트 데이터를 구축한다. 이 데이터셋은 큐시트를 포함하기 때문에 배경음악 식별 성능을 측정하는 데 활용할 수 있다. 또한 음악과 대사 트랙을 각각 제공함으로써 음악-대사 분리 성능을 측정하는 데에도 활용할 수 있다. 마지막으로 음악이 쓰인 구간인지 또는 음성이 쓰인 구간인지를 판단하는 음악-음성 구간검출 성능을 측정하는 데에도 활용할 수 있다.

III. 방송 배경음악 큐시트 데이터셋 구축

방송 배경음악 편집을 위한 원시 데이터를 활용하여 개발된 음악 식별기의 성능을 정확히 측정하기 위한 방송 배경음악 큐시트 데이터셋을 구축한다. 대한민국에서 방영된 5개의 드라마에서 임의로 총 60회차의 에피소드를 선정하였다. 총 오디오의 길이는 60시간 37분가량이며, 각 회차의 오디오 길이는 정규 방송 시간과 같이 한 시간 내외이다.

3-1 배경음악-대사 트랙별 오디오 데이터셋 구축

베가스 프로젝트에서 배경음악 트랙, 그리고 대사 및 현장 소음이 포함된 대사 트랙을 추출하고, 이 둘의 혼합 트랙을 추출한다. 실제 배경음악 제작 현장에서 배경음악을 대사와 혼합할 때 전체적으로 배경음악의 볼륨을 조절하는 마스터 볼륨 정보가 있다. 이는 배경음악이 드라마의 분위기를 살리면서도 한편으로는 배우의 대사를 방해하지 않고 대사가 명확하게 전달될 수 있도록 대사가 없는 부분의 배경음악은 볼륨을 높이고 대사가 포함된 부분은 일시적으로 볼륨을 낮추어준다. 대사 트랙 역시 배우의 대사가 없는 구간의 볼륨을 최소로 하여 불필요한 잡음이 포함되지 않도록 한다. 베가스 프로젝트에 포함된 이러한 마스터 볼륨 정보를 활용하여 렌더링함으로써 실제 방송에서 쓰이는 것과 유사한 오디오 특성을 갖도록 배경음악 트랙과 대사 트랙 데이터셋을 수집하였다. 그림 1에서 하나의 트랙 전반에 걸쳐 파란색 꺾은 선으로 나타난 것이 마스터 볼륨이다. 샘플링 주파수는 11,025Hz로 설정하고 모노 오디오로 렌더링하였다.

3-2 중복을 제거한 원곡 음원 데이터셋 구축

배경음악을 식별하기 위해서는 배경음악에 포함된 원본 음원으로부터 음악의 핑거프린트를 추출하여 데이터베이스에 저장하게 된다. 이러한 원곡 음원 데이터셋을 구성하기 위하여 다수의 원곡 음원을 다음과 같은 기준으로 중복이나 빠짐 없이 정리하였다.

방송 배경음악을 합성할 때는 발매된 음원을 그대로 사용하기도 하지만 작곡가가 작곡 시 활용했던 악기별 트랙 중 일부만 사용하기도 한다. 예를 들어, 주인공의 테마곡을 사용하되 장면 따라 선율이 추가 되는 악기만 사용하기도 하고, 보컬이나 드럼 트랙을 제거하기도 한다. 때로는 보컬을 제거한 반주만 사용하기도 한다. 이렇게 자유롭게 악기 트랙을 조합하기 위하여 현장에서는 그림 2와 같이 하나의 원곡 음원에 포함된 악기별 트랙을 각각 저장하여 보유하고 있다. 프로젝트에 포함된 이와 같은 악기 트랙별 음원을 모두 데이터베이스에 등록하게 되면 유사한 곡들로 인해 정제된 큐시트를 얻기가 힘들다. 실제로 저작권 등록을 하는 곡 역시 트랙별 음악을 발매하는 것이 아니라 모든 트랙이 혼합된 최종본을 발매하게 된다.

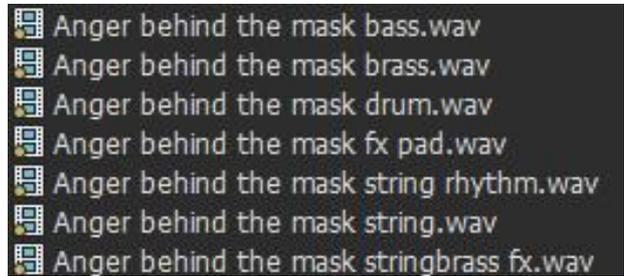


그림 2. 동일한 곡의 악기 트랙별 파일 예시

Fig. 2. An example of instrument tracks from the same song

배경음악 모니터링 시스템의 데이터베이스는 이렇게 발매된 곡을 기준으로 하여 구축하게 되므로 같은 제목의 트랙별 음원을 데이터베이스에 포함하지 않도록 하여 음악 식별 성능을 정확히 측정하고자 하였다.

방송 배경음악을 삽입하는 데 있어서 흔히 쓰이는 기법의 하나가 바로 피치 변환과 속도 변환이다. 피치 변환은 원곡 조성과 다른 조로 조를 옮기는 것이고, 속도 변환은 화면 전환 시점에 맞도록 원곡 음악의 길이를 사람의 귀로 눈치채지 못하게 늘이거나 줄이는 작업이다. 이러한 변환으로 인해 음악의 주파수 대역에 따른 에너지 값이 바뀌고 시간축 정보도 바뀌게 되어 음악의 특징정보 측면에서는 큰 변화가 발생하지만, 사람이 귀로 들으면 큰 차이를 느끼지 못해 원곡을 인식하는 데 어려움이 없는 경우가 대부분이다. 이러한 효과가 적용된 변환 음원 역시 원곡 음원 파일과 함께 존재되어 있는데 변환 음원은 제거하여 발매된 원곡만을 데이터베이스에 추가하도록 하였다.

원곡 음원의 전체가 아닌 일부만을 배경음악으로 삽입하는 경우 일부 구간만을 편집하여 오디오의 길이를 줄인 파일을 미리 별도의 음원으로 만들어 사용하는 예도 있다. 이런 음원이 데이터베이스에 추가되면 동일 곡에 대하여 중복곡이 발생하게 된다. 음원 파일 중에서 길이만 서로 다른 중복 음원을 제거하기 위하여 중복곡 제거 기법을 적용하였다.

방송의 배경음악과 달리 원곡 음원 데이터셋의 음원들은 다른 소리와 혼합되지 않은 음원이므로 기존의 음원 핑거프린팅 방법을 사용하면 중복 음원을 검출할 수 있다. 앞서 기술한 바와 같이 원곡 음원 데이터셋의 중복 음원들은 원곡 음악과 비교하면 일부 악기 트랙이 제외된 형태, 피치나 속도가 변환된 형태, 오디오의 길이를 줄인 형태 등의 왜곡이 가해진 음원들이다. 이러한 왜곡에 대하여 강인성을 가진 음원 식별 기술을 사용하면 데이터셋의 중복곡을 검출하여 제거할 수 있다.

한국저작권위원회에서는 위와 같은 변형을 포함하여 13개 변형 형태, 33개 세부 변형 파라미터를 적용한 왜곡 음원에 대하여 식별 기술의 성능을 검증하여 97% 이상의 식별 성능을 가진 기술에 대해 확인서를 발급하고 있는데[9], 본 논문에서는 한국저작권위원회의 성능 검증에서 99.9% 이상의 식별율을 보인 음원 식별기술[10]로 중복곡을 검출하여 제거하였다.

원곡 음원 데이터셋의 각 음원들에 대하여 해당 음원을 제외한 모든 음원과의 일치성을 각각 검사하여 90% 이상의 구간이 일치하는 음원이 존재하는 경우 중복곡으로 간주하였다.

일부 악기로 이루어진 음원, 피치나 속도 변화가 이루어진 음원, 음원 일부만을 포함하는 음원 등을 제거하여 최종적으로 중복곡 정제 후 드라마 5종에 포함된 원곡 음원은 총 723곡이다.

3-3 배경음악 식별 및 초기 큐시트 획득

음악 식별 시스템은 인식 대상 음원들의 핑거프린트를 추출하여 해당 음원 정보인 메타데이터와 함께 미리 데이터베이스에 저장한 후, 식별이 필요한 입력 음원의 핑거프린트로 데이터베이스를 검색하여 입력 콘텐츠의 메타 정보를 찾는다.

음악 식별 기술은 오디오 신호의 주파수 분석을 통해 얻은 스펙트로그램으로부터 음원의 핑거프린트를 추출하는 고전적인 방식인 랜드마크 기반 오디오 핑거프린트 기법[1], 차분해밍 거리 기반 이진 핑거프린트 기법[2]과 인공신경망을 통해 학습된 특징 벡터를 추출하는 방식[11]-[13]로 나눌 수 있다. 고전적인 핑거프린트 기반 식별 기법은 임계치를 설정하여 동일 음악인지 여부를 판단할 수 있음에 반해, 인공신경망으로 학습하여 얻은 핑거프린트의 경우 임계치를 설정하지 않고 단순히 가장 유사한 요소를 제시하는 방식으로, 방송 배경음악 식별 기술에서 필수적인 음악 유사도 판별의 기준 임계치를 얻기 어렵다. 이로 인해 음악이 없는 구간에서는 음악 식별 결과를 제시하지 않아야 하는 방송 배경음악 식별 기술에는 적용할 수 없는 방식이다.

본 논문에서는 다양한 음악 식별 기법 가운데 고전적인 방식[2]을 개량한 차분 기반 이진 해시 형태의 오디오 핑거프린트 기법[14]을 적용하여 음악을 식별한다. 정확한 큐시트를 얻기 위하여 대사 없이 배경음악만 포함된 음악 트랙 오디오를 입력으로 한다. 먼저 1초 간격의 프레임 단위로 나누어 배경음악을 식별한다. 전체 프레임에 대한 식별이 완료되면 프레임 식별 결과 중 오류로 보이는 것들을 제거하고, 프레임 결과들을 모아서 해당 방송 오디오에 대한 배경음악 구간 정보를 생성한다. 음악과 대사가 혼합된 실제 방송과 같은 혼합 트랙이 아닌 삽입된 배경음악만 포함하는 음악 트랙을 입력으로 하여 오디오 핑거프린트 기반 음악 식별 기법을 적용함으로써 상대적으로 정확한 초기 정답 큐시트를 얻을 수 있다. 식별 결과로 제시할 음원은 앞서 중복을 제거한 원곡 음원 데이터셋에 포함된 음원으로 한다.

3-4 정답 음악 구간 추출 및 큐시트 정제

음악만 존재하는 배경음악 트랙이라 하더라도 다양한 효과가 적용되었으므로 정확한 배경음악을 식별하지 못하는 경우가 발생할 수 있다. 정확한 큐시트를 제작하기 위해서는 음악이 포함된 모든 구간에서 음악 식별 결과가 빠짐없이 획득되

었는지를 확인해야 한다. 이를 위하여 배경음악만 포함된 음악 트랙에서 묵음구간이 아닌 신호가 포함된 구간 정보를 얻어 이를 정답 음악 구간으로 활용한다. 묵음이 아닌 구간 정보를 획득하는 데는 오디오 신호처리 라이브러리 librosa[15]의 librosa.effect.split 함수를 사용한다.

3.3절에서 획득된 배경음악 식별 결과와 정답 음악 구간을 비교하여 음악 구간임에도 불구하고 음악이 식별되지 않은 구간은 베가스 프로젝트를 참조하여 해당 곡의 아이디를 큐시트에 직접 기재한다. 이러한 경우는 주로 음원에 피치 변화나 속도 변환 효과가 적용된 경우가 대부분이다. 또한 식별은 되었으나 fade-in/out 효과로 인해 시작 또는 끝 시각 정보가 상이할 때는 정답 음악 구간 정보를 기준으로 식별된 음악의 시작 시각 또는 끝 시각을 수정하여 더욱 정확한 큐시트를 제작한다.

IV. 실험

3장에서 구축한 유사 방송 음악-대사 및 큐시트 데이터셋은 드라마 5종으로부터 수집된 60회차 분량의 오디오 트랙, 배경음악으로 사용된 원본 음원, 그리고 큐시트이다. 오디오는 대사가 중심이 되는 현장 녹음 트랙, 배경음악 트랙, 그리고 이 둘을 혼합한 혼합 트랙의 총 3종류로 구성된다. 오디오 데이터셋의 규모는 총 60.5시간씩 3트랙이다. 식별에 사용된 대상 음원 수는 드라마 5종에 포함된 중복을 제거한 원본 음악 723곡이다.

유사 방송 음악-대사 및 큐시트 데이터셋을 활용하여 배경음악 식별 성능을 측정해본다. 또한 배경음악과 대사가 각각의 트랙으로 되어 있으므로 음악-대사 분리 성능을 측정하는데도 사용해본다. 이때 분리된 음악 오디오를 입력으로 배경음악을 식별하여 성능에 어떤 영향을 미치는지도 함께 살펴본다. 마지막으로 큐시트에 포함된 배경음악 시작 및 끝 시각 정보를 활용해 구간 검출 성능을 측정해본다.

4-1 배경음악 식별 성능

정답 큐시트와 예측된 큐시트를 비교하여 방송 배경음악 모니터링을 위한 음악 식별 성능을 자동으로 측정하기 위하여 사운드 이벤트 검출 관련 라이브러리인 sed_eval[16]을 활용하였다. 방송 배경음악 큐시트를 자동으로 제작하는 과정에서 음악의 시작과 끝 시각을 정확하게 예측하는 것은 사용된 배경음악이 어느 정도의 길이로 사용되었는지에 대한 정보를 제공하는 데 필요하지만 배경음악을 삽입하는 과정에서 흔히 사용하는 fade-in/out 기법으로 인해 시작 또는 끝의 예측 시점이 정확하지 않을 수 있다. 또한 저작권료를 분배하는 과정에서 배경음악으로 사용된 정확한 길이보다 더욱 중요한 정보는 해당 방송 프로그램에 포함된 정확한 곡을 빠짐없이 식별하는 것이다.

표 1. 음원 분리 기법 적용 전후 음악 식별 성능
Table 1. The performance of music recognition according to applying music-speech separation technique

The performance of music recognition						
time-resolution	mixed audio			separated music audio		
	P	R	F1	P	R	F1
1 sec	0.9889	0.7492	0.8490	0.9885	0.8014	0.8825
5 sec	0.9888	0.7675	0.8611	0.9884	0.8135	0.8901
10 sec	0.9884	0.7806	0.8695	0.9882	0.8218	0.8952
15 sec	0.9884	0.7901	0.8756	0.9881	0.8280	0.8990

이러한 저작권료 분배라는 큐시트 제작 목적에 맞추어 시작 및 끝 이벤트 시점의 정확도를 측정하는 event-based metric을 사용하지 않고 일정 프레임 간의 정확도를 비교하는 segment-based metric을 사용하였다. 다양한 단위 길이(time-resolution) 내에서 곡이 검출되었는지를 확인하여 정확도를 측정하였고, 그 결과를 표 1에 나타내었다. 식별 기법은 앞서 서술한 초기 큐시트를 구하는 과정에서 사용한 핑거프린트 기법[14]를 적용하였다.

식별 성능을 살펴보면, 정확도(precision)는 0.99에 가까워 입력 오디오 단위 길이가 1초, 5초, 10초, 15초로 달라짐에 따라 미세하게 성능이 저하되긴 하지만 매우 우수한 성능을 보였다. 반면에 재현율(recall)은 1초일 때 0.7492에서 15초일 때는 0.7901까지 성능이 향상됨을 확인하였다. 이러한 재현율의 성능 향상으로 인해 최종 F1-score는 1초일 때 0.8490에서 15초일 때 0.8756까지 향상함을 확인하였다. 즉, 해당 기법의 방송 배경음악 식별 성능은 정확도는 0.99에 육박하여 예측된 음악은 매우 정확하게 식별하였지만, 재현율은 그에 비해 낮아 배경음악이 사용되었음에도 식별해내지 못하는 경우가 상대적으로 많다는 것을 확인하였다. 추후 재현율을 저하하는 요인을 분석한다면 방송 배경음악 식별 성능을 개선하는 데 크게 기여할 것으로 생각된다.

4-2 음악-대사 분리 기반 배경음악 식별 성능 및 음원 분리 성능

대사가 혼합된 방송 오디오에서 배경음악의 식별 성능을 높이는 한 가지 방법으로 음악과 대사가 혼재된 혼합 트랙에서 먼저 음악-대사 분리 기법을 적용하여 대사를 제거한 후 배경음악만 남도록 한 오디오 파일을 입력으로 동일한 기법을 활용해 배경음악을 식별해 보았다.

음악-대사 분리에 사용한 기법은 음악 음원 분리 기법 가운데 하나인 LaSAFT[17] 모델을 사용하였다.

표 2. 음악-대사 분리 성능
Table 2. The performance of music-speech separation

The performance of music-speech separation			
input	SDR	SIR	SAR
music	6.76	25.74	6.93
speech	7.84	11.93	10.81

LaSAFT 모델은 음악 음원 분리 기술 성능을 비교하는 가장 대표적인 데이터셋인 MUSDB18[18]을 기준으로 음원 분리 분야에서 원음 대비 왜곡의 정도를 나타내는 지표인 SDR (Source-to-Distortion Ratio)을 측정하였을 때 Demucs[19]에 비해 전체적으로 낮은 성능을 보이지만, 음악에서 가수의 목소리를 분리하는 데에는 훨씬 높은 성능을 보인다. 또한 보컬과 동일하게 사람의 목소리인 대사를 분리하는 데 LaSAFT[17]의 대사 분리 성능이 Demucs[19]보다 우수함을 보였다[20]. 본 논문에서는 대사 분리에 우수한 성능을 보인 LaSAFT[17]을 활용한 사전 학습된 음악-대사 분리 모델[20]을 적용하여 유사 방송 큐시트 데이터셋에서 대사 분리 전후 식별 성능의 변화를 관찰하고 그 결과를 표 1에 나타내었다. 사전 학습에는 음악-대사 분리 데이터셋[21]이 활용되었다. 이 데이터셋의 규모는 학습과 검증에 각각 12초 길이 1,823개 오디오를 활용한 약 12시간 규모임에 반해 본 데이터셋은 실제 방영분 길이와 동일한 1시간 내외 길이의 드라마 60회 분량으로 상대적으로 규모가 약 5배가량 크다.

대사를 분리하지 않은 혼합 트랙을 입력으로 한 배경음악 식별 성능과 대사를 분리한 오디오를 입력으로 한 성능을 비교하였다. 표 1에서와 같이 입력 오디오의 단위 길이가 짧은 경우, 즉 time-resolution이 1초인 경우, 분리 전 성능이 F1-score 기준 0.8490에서 분리 후 0.8825로 향상된 것을 볼 수 있다. 단위 길이가 긴 경우, 즉 time-resolution이 15초인 경우, 0.8756에서 0.8990으로 향상된 것에 비해 향상 폭이 큰 것을 알 수 있다. 즉, 대사가 분리된 오디오를 입력으로 하는 경우 입력 오디오 단위 길이에 따른 성능이 비교적 고르다는 것을 알 수 있다. 정확도의 경우 분리 전보다 분리 후 성능이 극히 미미한 수준으로 떨어지긴 하지만 여전히 0.99에 육박하는 우수한 성능을 보인다. 결국 재현율 기준 성능이 향상됨으로써 F1-score 기준 성능이 향상됨을 확인할 수 있다.

BAF 데이터셋[10]에서는 정확한 큐시트라 할지라도 5초의 오차범위가 존재한다고 보았는데 본 실험에서는 입력 단위 길이가 기준이 5초인 경우 음악 분리 전후 성능이 F1-score 기준 0.8611에서 0.8901로 향상되어 대사를 분리하는 과정이 방송 배경음악 식별 성능 향상에 도움이 됨을 알 수 있다.

다음으로 음악과 대사 음원을 분리하는 성능을 측정하였다. 음원 분리 분야에서 널리 쓰이는 평가 지표 라이브러리인 BSS eval[22]을 활용하여 분리 성능 지표인 SDR (source-to-distortion ratio)과 SIR (source-to-interference ratio), 그리고 SAR (source-to-artifacts ratio)를 측정하였다.

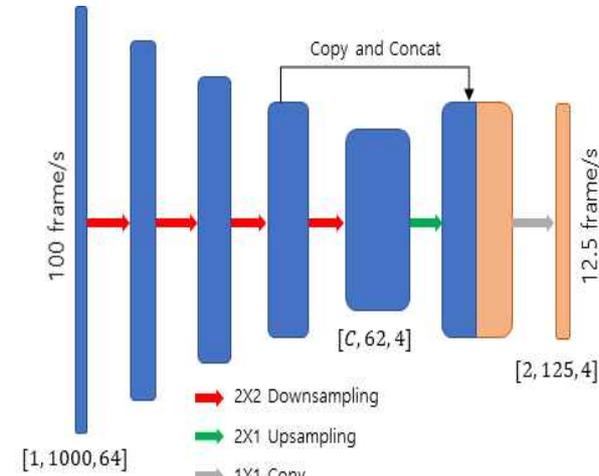


그림 3. 음악 구간 검출 네트워크 구조도
Fig. 3. The structure of music detection network

측정 시 mir_eval 툴박스[23]를 사용하였다. 음악-대사 분리 성능 측정 결과, 표 2와 같이 음악 오디오의 SDR은 6.76임을 알 수 있다. 이는 [21]에서 음악과 대사의 신호비가 -10dB인 경우 SDR이 6.84인 것과 유사한 수치임을 확인할 수 있다.

4-3 음악 구간 검출 성능

구축된 유사 방송 데이터셋에는 오디오에 포함된 곡의 아이디뿐만 아니라 구간 정보도 함께 포함되어 있으므로 대사와 음악이 혼합된 오디오에서 배경음악이 포함된 구간을 찾는 음악 구간검출 성능을 측정하는 데에도 직접적으로 활용할 수 있다. 이러한 음악 구간검출 기술은 실제 배경음악 모니터링 과정에서 정확한 방송 배경음악 모니터링 결과를 제공하기 위하여 음악 식별에 실패한 경우를 파악하는 데에 필요한 기술이다. 즉, 음악 구간검출을 통해 음악 구간으로 판별되었으나 음악 식별 결과를 얻지 못한 경우 해당 구간에 미정(Unknown)을 의미하는 특정 아이디를 부여함으로써 추후 관리자가 직접 개입하여 해당 음악 정보를 기재할 수 있도록 하는 데 활용할 수 있다.

본 논문에서는 컨볼루션 신경망 기반 방식[24]와 비대칭 U-Net 구조[25]를 음악 구간 검출 데이터에 맞게 개량한 기법 중에서 U-Net 기반 방식을 적용하여 성능을 측정해보았다. 개량된 네트워크는 그림 3에 나타난 것과 같이 U-Net[26]을 기반으로 영역분할이 아닌 분류 문제에 적합한 효율적인 구조로 변형되었다. 다양한 장르의 방송 프로그램으로부터 약 100시간의 방송 오디오를 수집 및 추출한 후 배경음악이 사용된 구간을 태깅하여 생성된 방송 음악 구간 데이터셋[27]로 사전 학습된 모델을 사용하였다.

구간검출 성능 측정은 음악 식별 성능 측정과 동일하게 사운드 이벤트 검출 관련 라이브러리인 sed_eval[16]을 활용하였다.

표 3. 음악 구간 검출 성능

Table 3. The performance of music detection

The performance of music detection			
input	Precision	Recall	F1-score
mixed audio	0.9742	0.9553	0.9644

큐시트에 표기된 곡명은 모두 'music'과 같은 동일 클래스로 변경한 후 일정 프레임 간의 정확도를 비교하는 segment-based metric을 사용하여 성능을 측정하였다. 정확한 구간 검출 성능 측정을 위하여 성능을 비교하는 단위 시간(time_resolution)은 1초로 설정하였다.

표 3과 같이 비대칭 U-Net 기법[25]를 활용한 음악 구간검출 기법의 정확도와 재현율은 각각 0.9742와 0.9553으로 0.95 이상의 높은 성능을 보이고 F1-score는 0.96 이상임을 보였다.

V. 결 론

본 논문에서는 실제 TV 방송의 배경음악 식별 성능을 정확하게 측정할 수 있는 유사 방송 음악-대사 및 큐시트 데이터셋을 구축하는 방법에 관하여 서술하였다. 유사 방송 음악-대사 및 큐시트 데이터셋은 각 방송 프로그램 회차별 오디오 파일, 방송에 포함된 배경음악 원본 음원, 그리고 배경음악의 곡 정보와 삽입된 시간 정보를 포함하는 큐시트로 구성된다. 오디오 파일은 방송 촬영 현장에서 녹음된 대사 트랙과 배경음악 트랙, 그리고 이 두 트랙의 혼합 트랙으로 이루어진다. 배경음악 트랙을 입력으로 기존의 음악 식별 기술을 적용하여 초기 큐시트를 획득하고, 음악 구간을 추출하여 이를 정제한 최종 정답 큐시트 데이터셋을 구축한다. 이렇게 구축된 데이터셋은 실제 방송 현장에서 녹음된 오디오와 배경음악 제작 프로덕션의 원시 데이터로부터 구축되어 실제 방송 오디오와 매우 흡사하면서도 사람이 직접 방송 오디오를 청취하여 작성한 큐시트 데이터가 가질 수 있는 한계점, 즉, 실제로 배경음악으로 삽입되었더라도 큐시트 목록에서 빠지는 경우가 존재하지 않아 큐시트의 정확도가 매우 높다.

본 논문에서 구축한 데이터셋은 방송 배경음악 모니터링을 위한 핵심 기술인 배경음악 식별 기법의 성능 측정뿐만 아니라 배경음악 모니터링의 요소 기술로 활용 가능한 음악-대사 분리 및 음악 구간 검출에 관한 기법의 성능 측정도 가능하다. 먼저 방송 배경음악을 식별하는 성능을 측정하여 다양한 입력 오디오 길이에 따른 배경음악 식별 성능을 제시하였다. 또한 음악-대사 분리 기법을 적용하면 배경음악 식별 성능이 개선됨을 함께 보였고, 음악 트랙과 대사 트랙 오디오가 각각 제공되어 음악-대사 분리 기법의 음원 분리 성능 자체를 측정하는 데에도 활용할 수 있음을 보였다. 마지막으로, 큐시트에 포함된 배경음악의 시작 및 끝 시각 정보를 활용하여 배경

음악이 삽입된 음악 구간을 검출하는 성능을 측정하는 데에도 활용할 수 있음을 보였다.

방송 배경음악 식별 성능 측정 결과를 살펴보면 음악-대사 분리 기법을 적용하여 배경음악만 포함하는 오디오를 입력으로 한 경우 F1-score 기준 약 0.9에 근접하는 성능을 보였다. 정확도는 0.98 이상으로 매우 높으나, 재현율은 최대 0.8280의 다소 낮은 성능을 보임을 확인하였다. 추후 식별에 실패한 경우의 배경음악과 대사 오디오를 함께 분석하여 실패의 주요 요인을 찾고, 이를 개선한 배경음악 식별 기법을 개발하고자 한다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2022년도 저작권기술 연구개발사업으로 수행되었음(과제명 : 딥러닝을 활용한 고속 음악 탐색 기술 개발, 과제번호 : CR202104004)

참고문헌

- [1] A. Wang, "An Industrial-Strength Audio Search Algorithm," in *Proceeding of the International Conference of Music Information Retrieval*, pp. 7-13, 2003.
- [2] J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System," in *Proceeding of the International Society for Music Information Retrieval Conference*, Paris, France, pp. 107-115, 2002.
- [3] Vegas. Available: <https://www.vegascreativesoftware.com/>.
- [4] Y.-N. Hung, C.-W. Wu, I. Orife, A. Hipple, W. Wolcott, and A. Lerch, "A large TV dataset for speech and music activity detection," *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2022, No. 21, pp. 1-12, September, 2022. <https://doi.org/10.1186/s13636-022-00253-8>
- [5] N. Schmidt, J. Pons, and M. Miron, "PodcastMix: A dataset for separating music and speech in podcasts," in *Proceeding of Interspeech*, Incheon, Korea, pp. 231-235, September, 2022.
- [6] C. Veaux, J. Yamagishi, and K. MacDonald, "Superseded-cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit," 2016.
- [7] S. Bazen, L. Bouvard, and J.-B. Zimmermann, "Jamendo: The Heartbeat of Free Music! Musicians and the Creative Commons," Aix-Marseille University, Technical Report, October, 2014.
- [8] G. C. Sebastià, Á. Ciurana, E. Molina, M. Miron, O. Meyers, J. Six, and X. Serra, "BAF: An Audio Fingerprinting Dataset For Broadcast Monitoring," in *Proceeding of the 23rd International Society for Music Information Retrieval Conference*, Bengaluru, India, December, 2022.
- [9] Korea Copyright Commission [Interten]. Available: <https://www.copyright.or.kr/business/tmis/performance/filtering/init.do?tabNm=techEval>
- [10] J. Park, J. Kim, H. Kim, "Feature Extraction and Search for Broadcasting Background Music Identification in A Large-Scale Music DB" in *Proceeding of the Korean Institute of Broadcast and Media Engineers*, pp. 483-484, July, 2020.
- [11] B. A. Arcas, B. Gfeller, R. Guo, K. Kilgour, S. Kumar, J. Lyon, J. Odell, M. Ritter, D. Roblek, M. Sharifi, and M. Velimirovic, "Now playing: Continuous low-power music recognition," in *Proceeding of NeurIPS 2017 Workshop on Machine Learning on the Phone and other Consumer Devices*, Long Beach, CA, 2017. <https://doi.org/10.48550/arXiv.1711.10958>
- [12] S. Chang, D. Lee, J. Park, H. Lim, K. Lee, K. Ko, and Y. Han, "Neural audio fingerprint for high-specific audio retrieval based on contrastive learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Ontario, Canada, pp. 3025-3029, 2021. <https://doi.org/10.1109/ICASSP39728.2021.9414337>
- [13] A. Báez-Suárez, N. Shah, J. A. Nolasco-Flores, S.-H. S. Huang, O. Gnawali, and W. Shi, "SAMAF: Sequence-to-sequence Autoencoder Model for Audio Fingerprinting," *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 16, No. 2, pp. 1-23, May, 2020. <https://doi.org/10.1145/3380828>
- [14] J. S. Seo, J. Kim, and H. Kim, "Audio fingerprint matching based on a power weight," *The Journal of the Acoustical Society of Korea*, Vol. 38, No. 6, pp. 716-723, 2019. <https://doi.org/10.7776/ASK.2019.38.6.716>
- [15] B. McFee, C. Raffel, D. Liang, D. P.W. Ellis, M. McVicar, E. Battenberg, and O. Nietok, "Librosa: Audio and music signal analysis in Python," in *Proceeding of Python in Science Conference*, pp. 18-25, 2015.
- [16] A. Mesaros, T. Heittola, and T. Virtanen, "Metrics for polyphonic sound event detection," *Applied Sciences*, Vol. 6, No. 6, pp. 1-17, May, 2016. <https://doi.org/10.3390/app6060162>
- [17] W. Choi, M. Kim, J. Chung, and S. Jung, "LaSAFT: Latent Source Attentive Frequency Transformation for Conditioned Source Separation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Ontario, Canada, pp. 171-175, June, 2021.

<https://doi.org/10.1109/ICASSP39728.2021.9413896>

- [18] Z. Rafii, A. Liutkus, F.-R. Stoter, S. I. Mimilakis, and R. Bittner, "MUSDB18 - a corpus for music separation," 2017.
- [19] A. Defossez, N. Usunier, L. Bottou, and F. Bach, "Music source separation in the waveform domain," *arXiv preprint arXiv:1911.13254*, 2019.
- [20] H. Kim, J. Kim, and J. Park, "Performance Analysis for Background Music Identification in TV Contents According to State-of-the-art Music Source Separation Methods," in *Proceeding of Korea Multimedia Society*, pp. 30-32, 2021.
- [21] H. Kim, W.-H. Heo, J. Kim, and J. Park, "Monaural Music-Speech Source Separation Based on Convolutional Neural Network for Background Music Identification in TV Shows," *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 45, No. 05, pp. 855-866, 2020.
<https://doi.org/10.7840/kics.2020.45.5.855>
- [22] E. Vincent, R. Gribonval, and C. Fevotte, "Performance Measurement in Blind Audio Source Separation," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 4, pp. 1462-1469, 2006.
<https://doi.org/10.1109/TSA.2005.858005>
- [23] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. P.W. Ellis, "mir_eval: A Transparent Implementation of Common MIR Metrics," in *Proceeding of International Society for Music Information Retrieval Conference*, Taipei, Taiwan, October, 2014.
- [24] B.-Y. Jang, W.-H. Heo, J. Kim, and O.-W. Kwon, "Music detection from broadcast contents using convolutional neural networks with a Mel-scale kernel," *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2019, No. 11, June, 2019.
<https://doi.org/10.1186/s13636-019-0155-y>
- [25] S. Lee, Y. Park, J. Park, G. Jang, and H. Kim, "Multi-target Learning on asymmetric U-Net for PNI boundary detection," in *Proceeding of International Conference on Big Data Applications and Services*, Jeju, Korea, November, 2021.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceeding of International Conference of Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, pp. 234-241, October, 2015. https://doi.org/10.1007/978-3-319-24574-4_28
- [27] H. Kim, J. Kim, and J. Park, "Music-Speech Separation based Background Music Identification in TV Programs," in *Proceedings of HCI Korea*, pp. 1158-1161, February, 2019.



김혜미(Hyemi Kim)

2006년 : 한국과학기술원 전기 및 전자공학부 (공학석사)

2006년~현 재: 한국전자통신연구원 통신미디어연구소 선임

2021년~현 재: 한국과학기술원 김재철AI대학원 박사과정

※ 관심분야 : 음악 정보 탐색(Music Information Retrieval), 딥러닝(Deep Learning)



김성우(Seongwoo Kim)

2017년~현 재 : 인하대학교 전자공학과 재학

※ 관심분야 : 오디오신호처리(Audio Signal Processing), 음악 정보 탐색(Music Information Retrieval)



이상원(Junghyun Kim)

2020년 : 경북대학교 전자공학과 (공학석사)

2020년~현 재: 경북대학교 전자공학과 박사과정

※ 관심분야 : 딥러닝(Deep Learning), 사운드 이벤트 검출(Sound Event Detection), 음성 분리(Speech Separation)



김정현(Junghyun Kim)

2001년 : 전남대학교 전산학과 (공학석사)

2001년~현 재: 한국전자통신연구원 통신미디어연구소 책임연구원

※ 관심분야 : 오디오 신호처리(Audio Signal Processing), 기계학습(Machine Learning)



박지현(Jihyun Park)

1999년 : 서강대학교 컴퓨터학과 (공학석사)

2005년 : 충남대학교 컴퓨터공학과 (공학박사)

1999년~현 재: 한국전자통신연구원 통신미디어연구소 책임연구원

※ 관심분야 : 저작권보호(Copyright Protection), 신호처리(Signal Processing), 기계학습(Machine Learning)



유원영(Wonyoung Yoo)

1998년 : 전북대학교 영상정보공학과 (공학석사)

2003년 : 전북대학교 전자공학과 (공학박사)

2001년~현 재: 한국전자통신연구원 통신미디어연구소 콘텐츠인식연구실장

※ 관심분야 : 저작권보호(Copyright Protection), 신호처리(Signal Processing)