논문 2025-2-3 http://dx.doi.org/10.29056/jsav.2025.06.03

노이즈 환경에서도 강인한 오디오 DNA 기반의 OTT 콘텐츠 인식 방법

박병찬*, 장세영*, 김석윤*, 김영모*†

A Robust Audio DNA-Based Method for OTT Content Recognition in Noisy Environment

Byeong-Chan Park*, Se-Young Jang*, Seok-Yoon Kim*, Young-Mo Kim*†

요 약

OTT 콘텐츠의 이용 확산과 함께 다양한 재생 환경에서의 저작권 보호 요구가 높아지고 있다. 특히, 백색 소음, 대화 소음 등 비정형 노이즈가 혼입된 상황에서도 원본 오디오의 식별이 가능한 기술이 중요해지고 있다. 본 논문에서는 이러한 환경에서 강인하게 인식될 수 있는 오디오 DNA 추출 및 인식 방법을 제안한다. Mel-spectrogram 기반의 특성 추출을 통해 coarse-fine 이중 단계의 특징을 생성하고, 다양한 파라미터 조합에 따른 인식 성능을 비교 평가하였다.

실험 결과, 최적화된 FFT 길이와 feature dimension 설정을 통해 높은 정확도를 달성하였으며, 제안한 방식은 OTT 환경에서의 실시간 콘텐츠 인식과 저작권 보호 시스템에 효과적으로 적용될 수 있음을 확인하였다.

Abstract

With the increasing use of OTT content, there is a growing demand for copyright protection across diverse playback environments. In particular, technologies capable of identifying original audio even under the presence of unstructured noise—such as white noise or conversational background noise—are gaining importance.

This paper proposes a robust method for extracting and recognizing audio DNA that remains reliable in such noisy conditions. By leveraging feature extraction based on MEL spectrograms, the proposed system generates a dual-stage fingerprint structure consisting of coarse and fine features. Recognition performance was evaluated under various parameter combinations.

Experimental results demonstrate that by optimizing FFT length and feature dimension, the proposed method achieves high accuracy. The results confirm that the approach can be effectively applied to real-time content authentication and copyright protection systems in OTT service environments.

한글키워드 : 오디오 DNA, 노이즈 환경, OTT 콘텐츠 인식, 오디오 핑거프린팅, 저작권 보호

keywords : Audio DNA, Noisy Environment, OTT Content Recognition, Audio Fingerprinting, Copyright protection

* 숭실대학교 컴퓨터학과
† 교신저자: 김영모(email: ymkim828@ssu.ac.kr)

접수일자: 2025.04.24. 심사완료: 2025.05.17.
게재확정: 2025.06.20.

1. 서 론

ICT환경과 OTT(Over-the-Top) 플랫폼의 급속한 성장으로 사용자들은 스마트폰, 테블릿, 노트북 등 다양한 디바이스를 통해 언제 어디서나 콘텐츠를 소비하게 되었다[1]. 이러한 다양한 환경의 콘텐츠 소비 환경은 사용 편의성을 높였지만, 동시에 불법 복제 및 유통과 같은 저작권 침해 문제와 이를 대응하기 위한 저작권 침해탐지 기술의 정확성을 낮추는 결과를 초래하고 있다. 특히, 다양한 재생 환경에서 발생할 수 있는 잡음, 압축 손실, 디지털 변형 등은 기존의 오디오 식별 기술이 정확하게 작동하지 못하도록 하는 주요 원인이 되었다.

기존 오디오 fingerprinting 기술은 원본과 완벽히 일치하는 클린 오디오를 기준으로 설계되어 있어, 실제 스트리밍 상황에서의 노이즈나 음질 저하를 견디는데 한계가 있다[2][3]. 이는 저작권 보호뿐만 아니라, 실시간 방송 모니터링, 콘텐츠 진위 확인, 그리고 사용자 맞춤형 콘텐츠 추천 등 다양한 응용 영역에서 심각한 신뢰도 문제를 야기할 수 있다[4][5].

따라서 본 논문에서는 오디오 신호가 다양한 종류의 노이즈에 노출된 상황에서도 원본 콘텐츠를 정확히 인식하고 구분할 수 있는, 노이즈 환경에 강인한 오디오 DNA 기반의 OTT 콘텐츠 인식 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 오디오 DNA 추출 기술 등을 기술한다. 3장에서는 본 논문에서 제안하는 노이즈 환경에 강인한 오디오 DNA 추출 및 인식 방법을 기술한다. 4장에서는 실험 및 결과를 보고 5장에서 결론으로 마무리한다.

2. 관련 연구

2.1 STFT 및 시간-주파수 변환 기반 오디오 분석

오디오 신호의 주요 특징을 추출하기 위한 기초적인 접근으로 단시간 푸리에 변환(STFT, Short-Time Fourier Transform)이 널리 사용된다. STFT는 시간-주파수 영역의 정보를 동시에 표현할 수 있어, 음악이나 음성 신호의 시간적인 변화를 반영한 분석에 적합하다[5-7]. 대부분의 오디오 인식 시스템에서는 입력 신호를 고정된 프레임 단위로 나눈 후, Hamming 또는 Hanning 윈도우를 적용하여 시간 축의 경계를 완화하고, 이를 통해 얻은 주파수 스펙트럼을 기반으로 특징을 추출한다[7].

2.2 MEL 필터와 인간 청각 모델 기반 특징 추출

STFT 이후에는 MEL 필터 뱅크(Mel-Scale Filter Bank)가 적용되며, 이는 인간의 청각 특성을 반영한 주파수 축 재배열 방식이다. MEL 스케일은 고주파보다 저주파에 더 민감한 인간의 인지 특성을 반영하므로, 불필요한 고주파 성분을 제거하면서도 신호의 주요 정보를 효과적으로 유지할 수 있다. 이 과정을 통해 생성된 Mel-Spectrogram은 fingerprint 추출의 기초 자료로 활용된다[4].

2.3 오디오 Fingerprint와 Coarse-Fine 구조

오디오 Fingerprint는 오디오 신호로부터 고유한 특징 값을 추출하여, 해당 오디오를 빠르게 식별하거나 검색할 수 있도록 하는 기술이다[3][8]. 일반적으로 fingerprint는 해시 형태의 이진 벡터(binary vector)로 표현되며, 기존 연구에서는 Shazam과 같이 peak-pairing 기반의 로컬 주파수 특징을 주로 활용해왔다.

그러나 최근에는 fingerprint를 coarse-fine 구조로 분리하는 연구가 진행되고 있다. Coarse fingerprint는 빠른 후보 검색을 위한 저차원의 이진 벡터로, 해시 인덱싱 구조에 적합하며, Fine fingerprint는 차분 기반 정규화를 통해 생성된 다중 클래스 벡터로, 정밀 매칭을 위한 시퀀스 정합 알고리즘(DTW, Cosine similarity 등)에 활용된다[9].

3. 노이즈 환경에 강인한 오디오 DNA 추출 및 인식 방법

3.1 개요

본 논문에서 제안하는 OTT 콘텐츠의 노이즈 환경의 강인한 오디오 DNA 추출 및 인식 방법으로 노이즈가 혼입된 오디오와 원본 오디오 간의 유사성을 효과적으로 비교하기 위해, 정확도 높은 DNA 추출 과정 및 구현된 추출 모듈로 구성되어 있으며, 그림 1과 같다.

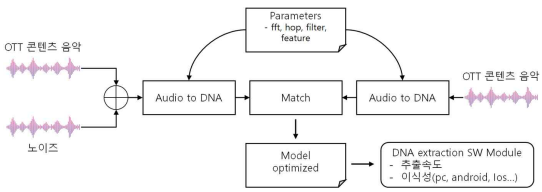

그림 1. 노이즈 환경에서도 강인한 오디오 매칭을 위한 시스템 구성도

Fig. 1. System Configuration for Robust Audio Matching in Noisy Environments

Audio to DNA 단계에서는 입력된 오디오로부터 식별력이 높은 DNA를 추출하는 과정이 수행된다. 이때 사용되는 주요 파라미터에는 FFT 길이, FFT 간격, 필터링 기법 등 다양한 요소가 포함되어있다. 추출된 DNA는 Match 단계에서 비교 대상 DNA와의 시퀀스 유사도 분석을 통해 일치 여부를 판단한다.

이러한 전체 과정을 기반으로 개발된 DNA extraction SW Module은 빠른 추출 속도와 높은 이식성을 보장하도록 설계되어, 다양한 플랫폼(pc, android, ios 등)에서도 사용 가능하다.

3.2 오디오 DNA 추출 및 매칭 과정

오디오 DNA 추출 과정은 그림 2와 같다.

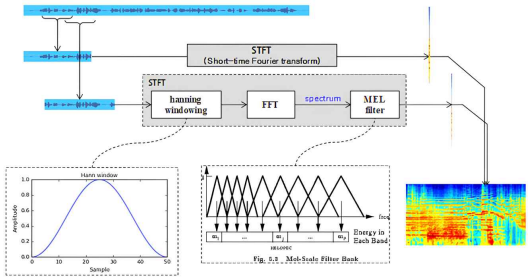

그림 2. STFT 및 MEL 필터 기반 오디오 특징 추출 과정

Fig. 2. STFT and MEL Filter-Based Audio Feature Extraction Process

입력 오디오는 시간 영역에서 일정 길이의 프레임으로 분할되며, 인접 프레임 간에는 중첩(overlap)을 두어 시간 해상도를 확보한다. 각 프레임에는 해밍 윈도우(Hamming Window)가 적용되어 경계 부분에서 발생할 수 있는 스펙트럼 왜곡을 완화한다. 이어서 짧은 시간 구간에 대한 푸리에 변환(STFT: Short-Time Fourier Transform)이 수행되어 시간-주파수 영역의 분포를 확보한다. SIFT는 식(1)을 기반으로 수행된다.

$$STFTx(t)(m,w) = \sum_{n=-\infty}^{\infty} x[n]\cdot w[n-m]\cdot e^{-jwn} \quad (1)$$

여기서 $x[n]$은 입력 오디오 신호, $w[n]$은 해밍 윈도우이며, 식(2)와 같이 정의된다.

$$w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \quad (2)$$

SIFT 결과로 얻어진 선형 스펙트럼에 MEL 필터 뱅크를 적용하여, 인가의 청각 특성을 반영한 MEL 스펙트로그램을 생성한다. 주파수 f를 MEL 주파수 m로 변환하는 과정은 식(3)에 의해 수행된다.

$$m = 2595 \cdot \log_{10}(1 + \frac{f}{700}) \quad (3)$$

MEL 필터링을 가진 Mel-spectrogram은 이후 특징 추출 단계의 입력으로 사용되며, 시간 및 주파수 축에서 지역적 구조를 반영한 이진 특징(binary features)을 추출하게 된다. 이 과정을 통해 corase fingerprint와 fine fingerprint의 이중 구조를 갖는 오디오 DNA가 형성되며, 그림 3과 같다.

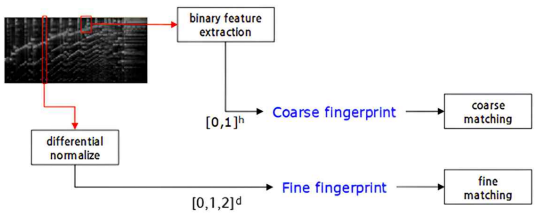

그림 3. 이중 Fingerprint 기반 오디오 인식 구조
Fig. 3. Audio Recognition Architecture Based on Dual Fingerprint Structure

Coarse fingerprint는 스펙트로그램의 이진 패턴을 기반으로 하여 생성되며, 이는 [0,1]^h 형태의 바이너리 시퀀스로 구성된다. 스펙트로그램의 세기 값에 대해 이진 임계값 처리를 통해 생성되며, 식(4)와 같이 표현된다.

$$f_c[i] = \begin{cases} 1 \text{ if } s_i \geq \theta_c \\ 0 \text{ otherwise} \end{cases} where\ f_c \in 0,1^h \quad (4)$$

여기서 s_i는 Mel-spectrogram의 에너지 값이며, θ_c는 임계값이다. 이 fingerprint는 주로 대용량의 데이터베이스에서 빠른 후보 검색(fast candidate retrieval)을 위한 근사 매칭(coarse matching)에 사용된다. 해시 기반 인덱싱 또는 트리 기반 구조와 결합하여 연산 부하를 줄이면서도 높은 검색 속도를 확보할 수 있다.

Fine fingerprint는 coarse fingerprint로 추려진 후보들에 대해 보다 정밀한 비교를 수행하기 위해 활용되며, 식(5)에 의해 정의된다.

$$f_i[i] = \left[\frac{s_{i+1} - s_i}{\Delta}\right] where\ f_f \in 0,1,3^d \quad (5)$$

여기서 Δ는 스케일링 파라미터이다. 이 fingerprint는 정밀한 시퀀스 매칭(fine matching)을 위한 입력으로 사용되며, 시퀀스 정합도는 코사인 유사도인 식(6) 또는 해밍거리인 식(7)과 같은 거리 기반 알고리즘을 통해 측정된다.

$$sim(f_1, f_2) = \frac{f_1 \cdot f_2}{\|f_1\| \cdot \|f_2\|} \quad (6)$$

$$d_H = (f_1, f_2) = \sum_{i=1}^{n} 1_{f_1[i] \neq f_2[i]} \quad (7)$$

이와 같은 이중 구조의 fingerprint 설계는 연산 효율성과 정확도 간의 trade off를 효과적으로 해결하며, 다양한 노이즈 조건 하에서도 강인한 오디오 식별 성능을 보장한다. 또한 coarse 단계에서 빠른 필터링이 가능하므로 대규모 OTT 콘텐츠 데이터베이스에서도 실시간 적용이 가능하다.

4. 실험 및 결과

4.1 실험 환경 및 파라미터 설정

본 논문에서 제안하는 OTT 콘텐츠의 노이즈 환경에 강인한 오디오 DNA 추출 및 인식 방법을 실험 및 검증하기 위하여 표 1과 같은 실험 환경을 구성하였다.

표 1. 실험 환경
Table 1. Experiment environment

	Environment
CPU	Intel Core 12900K
GPU	NVIDEA Geforce RTX 4090

실험에 사용된 데이터는 총 649개의 오디오 샘플로, 각각의 샘플은 약 20초 길이의 22kHz mono WAV 형식으로 구성되었다. 실험용 데이터는 reference-query 쌍 형태로 구성되며, 그림 4와 같다.

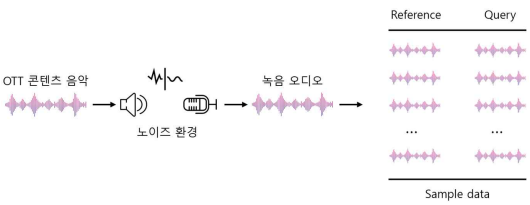

그림 4. 실험용 reference-query 샘플 생성 과정
Fig. 4. Generation Process of Reference-Query Samples for Experiments

Reference: 원본 음원 (다양한 장르 및 음색 포함)

Query: 동일 음원을 스피커로 출력 후, 다양한 노이즈 환경에서 마이크로 녹음하여 생성한 변형 음원 오디오 DNA의 성능에 영향을 줄 수 있는 주요 파라미터는 표 2와 같이 설정하였다.

실험에서는 위 변수들을 다양한 조합으로 구성하여, 각 설정 값이 분별력에 미치는 영향을 관찰하였다.

4.2 성능 평가 방법 및 실험 케이스

성능 평가는 DNA fingerprint 간 시퀀스 유사도를 기준으로 한다. 유사도 측정은 다음 두 항목으로 나뉜다.

Intra 유사도: 동일한 오디오(reference-query 쌍) 간의 fingerprint 유사도를 측정한다. 이는 동일한 콘텐츠가 시간이나 인코딩 방식 등 일부 조건만 다른 경우에도 얼마나 일관된 fingerprint를 생성하는지를 평가한다. 높은 Intra 유사도는 정합성(consistency)이 높음을 의미한다.

Inter 유사도: 서로 다른 오디오 간 fingerprint 유사도를 측정한다. 이는 상이한 콘텐츠 간 fingerprint가 얼마나 명확히 구분되는지를 보여주며, 낮은 Inter 유사도는 분별력(discriminability)이 높음을 의미한다.

이 두 값을 기반으로, 분별력(Discriminability)은 다음과 같은 식(8)으로 정의된다.

표 2. 오디오 DNA 추출을 위한 주요 파라미터 설정
Table 2. Key Parameter Settings and in Audio DNA Extraction

Parameter	설명	효과
FFT Length	FFT에 입력될 PCM 길이	길수록 주파수 정밀도 ↑, 시간 정밀도 ↓
FFT Hop Length	프레임 간 간격(overlap 조절)	짧을수록 시간 정밀도 ↑, bitrate ↑
Feature Dimension(Coarse)	Coarse fingerprint 벡터 차원 수	높을수록 분별력 ↑, 낮을수록 bitrate ↓
Feature Dimension (Fine)	Fine fingerprint 벡터 차원 수	높을수록 정밀매칭 정확도 ↑

$$Discriminability = mean(sim_{intra}) - mean(sim_{inter}) \quad (8)$$

각 실험은 다양한 파라미터 조합을 통해 진행되었으며, 주요 실험 케이스는 표 3과 같다.

4.3 결과 및 분석

- T00

실험 케이스 T00에서는 FFT 길이만을 변수로 설정하여 2048과 4096의 두 가지 설정에서 coarse 및 fine fingerprint 각각의 intra/inter 유사도 차이를 측정하였으며, 표 4와 같다.

또한, coarse와 fine fingerprint의 각 조건에서 intra/inter 유사도 시계열 분포를 나타내었으며, 그림 5와 같다.

그림 5의 상단 두 그래프는 FFT 길이 2048일 때, 하단은 4096일 때의 결과를 보여준다. 각각의 그래프에서 파란색은 intra(동일 오디오 간), 주황색은 inter(다른 오디오 간) 유사도를 나타낸다. 분별력은 intra 값이 낮고, inter 값이 높을수록 우수하다.

표 4의 실험 결과 4096 설정의 coarse와 fine fingerprint 모두에서 inter/intra 비율이 약 6% 이상 증가하며 분별력이 더 우수한 것으로 나타났다.

이는 주파수 해상도가 향상됨에 따라 더 정교한 특징 추출이 가능해졌기 때문으로 분석된다. 따라서 OTT 환경에서의 정확한 콘텐츠 인식을 위해서는 FFT 길이를 충분히 확보하는 것이 효과적임을 확인할 수 있다.

- T01

실험 T01에서는 FFT 길이를 고정(4096)한 상태에서 hop length(프레임 간 간격)만을 변수로 설정하여, 882와 1470 두 값에서 coarse 및 fine fingerprint의 분별력 지표를 비교하였다. Hop length는 시간 해상도와 fingerprint의 밀도에 영향을 주는 중요한 파라미터이며, 표 5와 같다.

분석 결과, hop length가 짧은 882 설정이 coarse와 fine 모두에서 inter/intra 비율이 약 1% 가량 더 우수한 것으로 나타났다. 이는 더 자주 프레임을 추출함으로써 시간 해상도가 향상되고, 결과적으로 더 정밀한 fingerprint가 생성되었기

표 3. FFT 및 Fingerprint 파라미터 조합별 실험 케이스
Table 3. Experimental Cases for Combinations of FFT and Fingerprint Parameters

ID	FFT Length	Hop Length	Coarse Dim	Fine Dim
T00	{2048, 4096}	882	128	128
T01	4096	{1470, 882}	128	128
T02	4096	882	{64, 128}	128
T03	4096	882	64	{64, 128}

표 4. FFT 길이에 따른 coarse/fine 분별력 정량 비교 (실험 T00)
Table 4. Quantitative Comparison of Coarse/Fine Discriminative Power According to FFT Length (T00)

FFT Size	Coarse			Fine		
	Intra	Inter	Intra/Inter	Intra	Inter	Intra/Inter
2048	23.41	29.10	1.257	21.25	26.50	1.261
4096	22.26	29.13	1.332	19.46	25.64	1.338

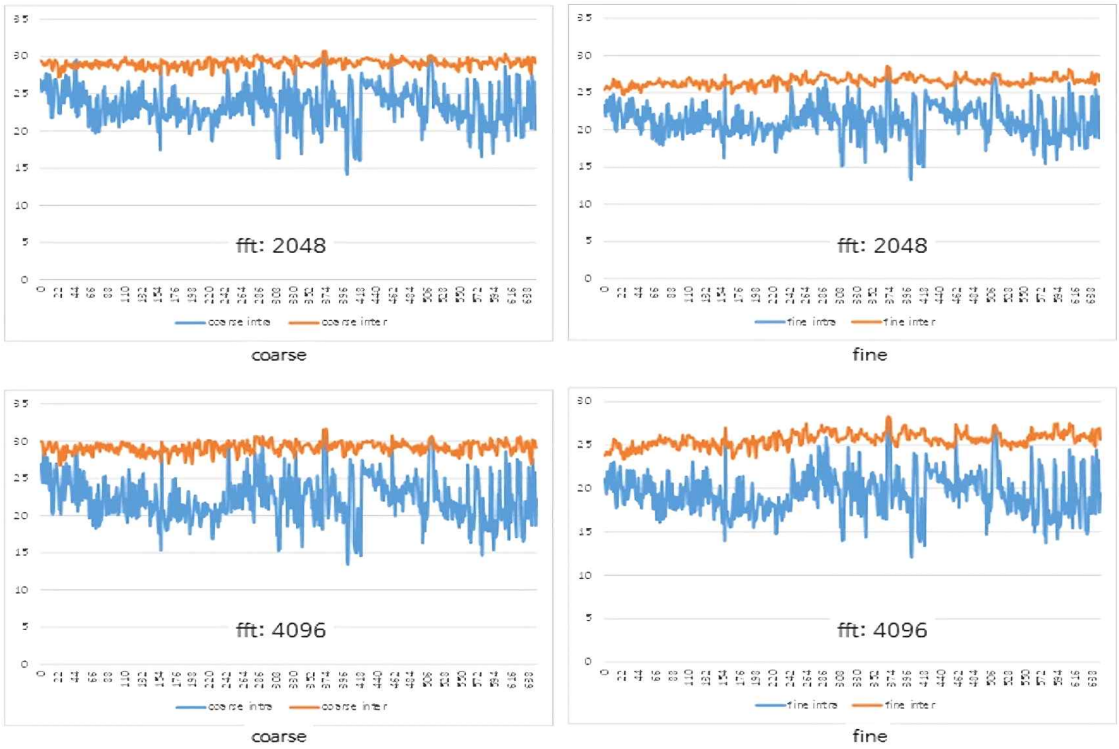

그림 5. FFT 길이에 따른 coarse 및 fine fingerprint의 intra/inter 유사도 비교 결과 (실험 T00)
Fig. 5. Comparison Results of Intra/Inter Similarity for Coarse and Fine Fingerprints According to FFT Length

표 5. Hop Length에 따른 coarse/fine 분별력 정량 비교 (실험 T01)
Table 5. Quantitative Comparison of Coarse/Fine Discriminative Power According to Hop Length (T01)

Hop Length	Coarse			Fine		
	Intra	Inter	Intra/Inter	Intra	Inter	Intra/Inter
882	22.25801	29.13482	1.331695	19.4586	25.63156	1.337812
1470	22.42108	29.13616	1.320801	19.5983	25.6275	1.326934

때문으로 해석된다.

따라서 실시간 콘텐츠 인식 정확도를 높이기 위해서는, 일정 수준 이하의 hop length 유지가 바람직하다. 다만, 짧은 hop length는 bitrate 증가로 이어지므로, 시스템 자원과 처리 속도에 따라 최적의 균형이 요구된다.

- T02

실험 T02는 coarse fingerprint의 차원 수(coarse dimension)를 변수로 하여, 64와 128 두 가지 설정에서의 분별력 성능을 비교하였다. 특징 벡터의 차원 수는 fingerprint의 정보량과 분별력에 직접적인 영향을 미치는 중요한 파라미터이며, 표 6과 같다.

표 6. Coarse Feature Dimension에 따른 분별력 비교 (실험 T02)
Table 6. Comparison of Discriminative Power According to Coarse Feature Dimension (T02)

Dim	Coarse		
	Intra	Inter	Intra/Inter
64	21.72236	29.7282	1.395886
128	22.25801	29.13482	1.331695

표 7. Fine Feature Dimension에 따른 분별력 비교 (실험 T03)
Table 7. Comparison of Discriminative Power According to Fine Feature Dimension (T03)

Dim	Fine		
	Intra	Inter	Intra/Inter
64	19.45858	235.63156	1.337812
128	18.66766	25.77689	1.40604

실험 결과, coarse dimension이 64일 때의 inter/intra 비율이 약 1.396으로, 128일 때보다 약 5% 향상된 분별력을 보였다. 이는 차원 수를 낮춤으로써 오히려 fingerprint 간 간섭(inter-pair similarity)이 줄어들고, 분리 가능성이 향상된 것으로 분석된다.

하지만 차원이 낮을수록 fingerprint의 표현력은 제한될 수 있으므로, 이 결과는 coarse 단계에서 빠른 후보 검색을 위한 fingerprint 설계에 적절한 기준으로 활용될 수 있다.

- T03

실험 T03은 fine fingerprint의 차원 수(fine dimension)를 변수로 설정하여, 64와 128 설정 간의 분별력 차이를 비교하였다. fine fingerprint는 정밀 매칭 단계에서 오디오 간 유사도를 평가하는 데 사용되므로, 특징 벡터의 정밀도가 인식 정확도에 중요한 영향을 미치며, 표 7과 같다.

실험 결과, fine dimension이 128인 경우 inter/intra 비율이 약 1.406으로, 64 대비 약 5% 향상된 분별력을 보였다. 이는 차원이 증가함에 따라 오디오의 세부 정보를 보다 정교하게 표현할 수 있게 되어, fine matching 단계의 정밀도가 개선된 결과로 해석된다.

다만, dimension 증가에 따른 메모리 사용량 및 연산량 증가를 고려할 때, 실제 시스템 적용 시 성능과 자원 간의 균형 조절이 필요하다.

- 종합 분석 및 최적 파라미터 설정

앞선 T00~T03 실험 결과를 종합적으로 비교한 결과, 인식 정확도와 효율성의 균형을 고려할 때 가장 효과적인 파라미터 조합은 표 8과 같다.

표 8. 실험 결과에 따른 최적 파라미터 설정
Table 8. Optimal Parameter Settings Based on Experimental Results

FFT Length	FFT Hop	Coarse Dimenesion	Fine Dimension
4096	1470	64	128

이 설정은 높은 주파수 정밀도(4096-point FFT)와 적절한 시간 해상도 유지(1470 hop length), 낮은 coarse dimension으로 인한 빠른 후보 필터링, 높은 fine dimension으로 인한 정밀 매칭 성능을 동시에 만족시킨다.

최적 파라미터 설정을 기준으로, coarse 및 fine fingerprint의 intra 유사도(시퀀스 거리)를 시계열로 비교하였으며 그림 6과 같다.

coarse는 64bit, fine은 256bit(128차원 × 2bit) 구조로 설계되었으며, 두 fingerprint 간의 유사도 차이를 통해 fine matching 방식이 보다 세밀한 구분 능력을 제공함을 확인할 수 있다.

이 결과를 통해, coarse 단계에서 빠른 후보 추출을 수행한 뒤, fine 단계에서 정밀한 최종 판별을 수행하는 coarse-to-fine 매칭 구조의 효과성이 입증되었다. 제안한 방식은 OTT 콘텐츠 환경과 같이 노이즈가 혼입된 비정형 오디오 상황에서도 실시간 식별 및 인증이 가능한 구조로서, 저작권 보호 시스템 적용에 매우 적합함을 실험적으로 확인하였다.

그림 6. 최적 파라미터 설정 기반 coarse/fine intra 유사도 시퀀스 비교

Fig. 6. Comparison of Coarse/Fine Intra Similarity Sequences Based on Optimal Parameter Settings

5. 결론

본 논문에서는 노이즈 환경에서도 강인한 오디오 DNA 추출 및 인식 기법을 제안하였다. 제안된 방법은 STFT 및 MEL 필터 기반의 전처리를 통해 시간-주파수 도메인에서 오디오의 핵심 특징을 추출하며, coarse-fine 이중 구조의 fingerprint를 생성하여 빠른 검색과 정밀한 매칭을 동시에 구현한다.

실험에서는 FFT 길이, hop 길이, coarse/fine 차원 수 등 다양한 파라미터에 따른 분별력을 측정하였고, 그 결과 FFT 4096, hop 1470, coarse dimension 64, fine dimension 128 조합이 가장 우수한 성능을 나타냈다. coarse fingerprint는 빠른 후보 필터링을, fine fingerprint는 2bit 정밀도를 기반으로 한 정밀 매칭을 통해 시스템 전반의 효율성과 정확도를 동시에 확보하였다.

제안된 방법은 OTT 환경에서의 실시간 콘텐츠 인증, 저작권 보호, 콘텐츠 추적 시스템 등에 효과적으로 적용 가능하다.

첫째, 다양한 노이즈 환경(백색소음, 실내 대화, 거리 소리 등)에서도 높은 Intra 유사도와 낮은 Inter 유사도를 유지하여 강인한 식별 성능을 보였다.

둘째, coarse-fine 구조를 통해 실시간 처리가 가능하며, 대규모 데이터셋에서도 검색 속도와 정확도를 모두 확보할 수 있다.

셋째, 실험 결과는 OTT 콘텐츠에서의 압축, 인코딩, 리샘플링과 같은 실제 환경 변형에 대해서도 일관된 fingerprint 생성을 확인하였다.

이러한 점에서 제안된 시스템은 실시간 콘텐츠 인증, 저작권 보호, 콘텐츠 무단 복제 탐지 등 다양한 OTT 및 스트리밍 기반 서비스에 실질적으로 적용 가능하다.

종합적으로 상용화 가능성이 높은 오디오 DNA 기반 콘텐츠 식별 기술을 제시하며, 향후 다양한 스트리밍 플랫폼과의 연동을 통해 상용화 가능성이 높다.

이 논문은 2025년도 문화체육관광부 및 한국콘텐츠진흥원의 재원으로 SW저작권기술(+법) 융합인재양성사업의 지원을 받아 수행된 연구임(과제명 : OTT 콘텐츠 저작권 보호 기술개발 및 적용을 위한 저작권기술 융합인재양성, 과제번호 : RS-2023-00225267)

참 고 문 헌

[1] Wooseop Lee, Seyoung Jang, Injae Yoo, Byungchan Park, Sunhee Shin, Seokyun Kim, and Youngmo Kim, “Illegal Streaming Video Recognition Method Based on Ending Credits for OTT Content Identification”, Journal of the Korea Software Appraisal and Valuation Society, Vol. 20, No. 4, pp. 225 - 230, 2024. DOI: http://dx.doi.org/10.29056/jsav.2024.12.23

[2] Yoohyun Son and Junyoung Heo, “Audio Fingerprint Generation Based on Multi-Time Segment Peaks”, Journal of the Korea Information Processing Society: Software and Data Engineering, Vol. 14, No. 1, pp. 48 - 52, 2025. DOI: https://doi.org/10.3745/TKIPS.2025.14.1.48

[3] P. Cano, E. Batlle, T. Kalker, J. Haitsma, “A Review of Algorithms for Audio Fingerprinting”, IEEE Workshop on Multimedia Signal Processing, 2002. DOI: https://doi.org/10.1109/MMSP.2002.1203274

[4] M. Huzaifah, “Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks”, arXiv preprint, arXiv:1706.07156, 2017. DOI: https://doi.org/10.48550/arXiv.1706.07156

[5] H. Jeon, Y. Jung, S. Lee, Y. Jung, “Area-Efficient Short-Time Fourier Transform Processor for Time - Frequency Analysis of Non-Stationary Signals”, Applied Sciences, vol. 10, no. 20, pp. 7208, 2020. DOI: https://doi.org/10.3390/app10207208

[6] S. D. Voran, “Why some audio signal short-time Fourier transform coefficients have nonuniform phase distributions”, arXiv preprint, arXiv:2409.08981, 2024. DOI: https://arxiv.org/abs/2409.08981

[7] A. Marafioti, N. Holighaus, P. Majdak, “Time-Frequency Phase Retrieval for Audio -- The Effect of Transform Parameters”, IEEE Transactions on Signal Processing, Vol. 69, pp. 3585-3596, 2024. DOI: https://doi.org/10.1109/TSP.2021.3088581

[8] J. Wang, Y. Zhang, J. Wang, “Sequence-to-sequence Autoencoder Model for Audio Fingerprinting”, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 16, no. 2, pp. 1 - 24, 2020. DOI: https://doi.org/10.1145/3380828

[9] Chang, S., Lee, D., Park, J., Lim, H., Lee, K., Ko, K., & Han, Y, “Neural Audio Fingerprint for High-specific Audio Retrieval based on Contrastive Learning”, arXiv preprint arXiv:2010.11910. 2021. DOI: https://doi.org/10.48550/arXiv.2010.11910

저 자 소 개

박병찬(Byeong-Chan Park)

2015.2 학점은행제 졸업
2018.2 숭실대학교 컴퓨터학과 석사
2023.8 숭실대학교 컴퓨터학과 박사
2023.9-현재 숭실대학교 초빙교수
<주관심분야> 저작권 보호 및 이용활성화

장세영(Se-Young Jang)

2018.2 평생교육원 학점은행 졸업
2021.6 숭실대학교 컴퓨터학과 석사
2023.2-현재 숭실대학교 컴퓨터학과 박사 과정
<주관심분야> 저작권 보호 및 이용활성화

김석윤(Seok-Yoon Kim)

1980.2 서울대학교 전기전자 졸업
1990.2 University of Taxas at Austin Dept. of ECE 석사
1993.2 University of Taxas at Austin Dept. of ECE 박사
1982-1987 ETRI 연구원
1993-1995 모토로라 책임 연구원
1995-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화

김영모(Young-Mo Kim)

2003.2 대전대학교 컴퓨터공학과 졸업
2005.2 대전대학교 컴퓨터공학과 석사
2011.2 대전대학교 컴퓨터공학과 박사
2012-현재 : 숭실대학교 교수
<주관심분야> 저작권 보호 및 이용활성화