

# 미디어 접근편의성 향상을 위한 음향 데이터 삽입 및 색인 기술 동향

Trends in Acoustic Data Hiding and Tagging Technologies for  
Enhancement of Media Accessibility

성종모 (J. Sung, jmseong@etri.re.kr)    실감 AV 연구그룹 책임연구원  
백승권 (S. Beack, skbeack@etri.re.kr)    실감 AV 연구그룹 책임연구원  
이미숙 (M. Lee, lms@etri.re.kr)        실감 AV 연구그룹 책임연구원  
이태진 (T. Lee, tjlee@etri.re.kr)        실감 AV 연구그룹 책임연구원

\* 본 연구는 2017년도 ETRI 정부출연금사업(기관고유사업)의 일환으로 수행되었음[15ZR1200, 미디어 접근편의성 향상을 위한 Active Audioprint 기술 개발].

오디오 워터마크, 음향 데이터 전송 및 오디오 핑거프린트 등으로 대표되는 음향 데이터 삽입 및 색인 기술은 최근 다양한 미디어 활용 인프라의 보급과 새로운 형태의 미디어 생태계가 등장함에 따라 중요성이 더욱 커지고 있으며, 콘텐츠 제어 및 식별을 비롯한 다양한 응용 서비스의 기반 기술로 활용될 수 있다. 본고에서는 음향 신호 기반 데이터 삽입 및 색인 기술 개발 현황과 관련 서비스 동향에 대해서 소개한다.



본 저작물은 공공누리 제4유형  
출처표시+상업적이용금지+변경금지 조건에 따라 이용할 수 있습니다.

2017  
Electronics and  
Telecommunications  
Trends

멀티미디어 & 방송기술 특집

- I. 머리말
- II. 음향 데이터 삽입 기술 동향
- III. 음향 데이터 색인 기술 동향
- IV. 음향 데이터 삽입 및 색인 서비스 동향
- V. 맺음말

## I. 머리말

최근 초고속 인터넷망 및 스마트 기기를 비롯한 다양한 미디어 활용 인프라의 보급과 YouTube, Facebook, SnapChat 등 새로운 형태의 미디어 생태계가 보편화됨에 따라 이용자들은 과거보다 훨씬 다양한 방식으로 대규모의 콘텐츠를 즐길 수 있게 되었다. 미디어 생태계에서 콘텐츠는 일반적인 공산품과 마찬가지로 생산, 유통 및 소비로 이루어진 생명주기 단계를 거치게 된다. 각 단계에서 생산자는 콘텐츠에 대한 제어 권한을 갖기 원하고, 유통업자는 수익 창출을 위한 부가 가치 제공 기회와 콘텐츠 사용에 대한 효과적인 모니터링 권한을 필요로 하며, 소비자에게는 콘텐츠에 대한 편리한 접근 권한과 제어 기능을 제공해야 한다. 이러한 콘텐츠의 효율적인 관리를 위해서는 해당 콘텐츠에 대한 제어 및 식별 과정이 필수적이다. 종래에는 콘텐츠 식별, 저작권 보호, 복사 방지, 뷰어 제어 및 포렌식 등의 DRM(Digital rights management) 관점에서 주로 접근하였으나 최근에는 소비자에게 새로운 형태의 부가 서비스를 제공하기 위한 시도도 함께 이루어지고 있다.

콘텐츠 제어 및 식별 기술에는 제어 정보를 콘텐츠에 직접 삽입하는 디지털 워터마크와 콘텐츠 분석을 통해 내재된 고유한 정보를 추출하는 디지털 핑거프린트가 있으며, 유사한 목적을 위해 서로 다른 기술적 접근 방법을 채용하고 있다. 일반적으로 워터마크 기술은 콘텐츠에 대한 개별 자산 추적, 제작자 식별 및 콘텐츠의 합법적인 취득 여부를 판단하는데 유용한 반면, 핑거프린트 기술은 콘텐츠 모니터링, 저작권 제어, 메타데이터 연계 등의 응용에 적합한 것으로 알려져 있다. 두 기술 모두 이미지, 비디오, 오디오, 텍스트 등 다양한 콘텐츠 유형에 모두 적용할 수 있지만, 본고에서는 오디오와 같은 음향 신호를 대상으로 하는 경우에 한정하여 관련 기술들을 다루기로 한다.

최근에는 콘텐츠에 포함된 음향 신호를 매개로 하여

〈표 1〉 오디오 기반 콘텐츠 제어 및 식별 기술 비교

방식	오디오 핑거프린트	오디오 워터마크	ADT
기술적 목표	오디오 콘텐츠와 메타데이터 연동	오디오 신호에 소량의 데이터 은닉	오디오 신호를 통한 데이터 통신
원본 콘텐츠	유지	변경	변경
데이터 전송률	-	낮음	높음
데이터베이스	필요	불필요	불필요
사본 식별	불가	가능	가능
포맷 종속성	낮음	높음	높음
음향 채널 강인성	높음	낮음	높음
주요 응용분야	콘텐츠 모니터링 저작권 제어 메타데이터 연계 음악 검색	개별 자산 추적 제작자 식별 저작권 보호 포렌식	콘텐츠 연계 및 부가 서비스 (광고 등)

생산자/유통업자와 소비자 간의 정보 전달 수단을 제공하는 음향 데이터 전송(ADT: Acoustic data transmission) 기술이 개발되었다. ADT는 스피커와 마이크로 구성된 음향 채널을 통해 음향 신호를 전달 대체로 하는 데이터 통신 기술로, 이를 활용한 다양한 서비스 발굴이 활발하게 이루어지고 있다.

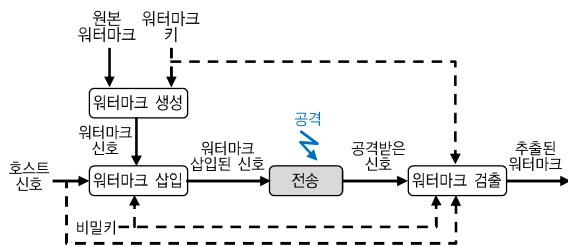
〈표 1〉은 앞서 설명한 오디오 신호를 기반으로 한 콘텐츠 제어 및 식별 기술들에 대한 특징을 비교한 것이다.

오디오 워터마크와 ADT 기술은 호스트 신호에 데이터를 삽입하는 공통점으로부터 음향 데이터 삽입 기술에 속하며, 오디오 핑거프린트 기술은 호스트 신호를 변경하지 않고 고유한 식별 정보를 추출한다는 면에서 음향 데이터 색인 기술로 구분한다. 본고에서는 사용자 중심의 콘텐츠에 대한 접근 편의성 향상 측면에서 음향 신호 기반 데이터 삽입 및 색인 기술 개발 현황과 서비스 동향에 관하여 기술한다.

## II. 음향 데이터 삽입 기술 동향

### 1. 오디오 워터마크 기술

일반적으로 인간의 청각 시스템은 시각 시스템에 비해 더 민감한 것으로 알려져 있다. 동일한 시간 간격에서 오디오 신호를 표현하는 샘플의 수가 영상 신호에 비



(그림 1) 오디오 워터마크 시스템 구조

해 훨씬 적기 때문에, 청각적으로 왜곡을 주지 않으면서 워터마크의 검출 성능을 높이는 것은 매우 어려운 주제로 알려져 있다[1]. 오디오 워터마크 기술은 용도에 따라 강성 워터마크와 연성 워터마크로 분류할 수 있다 [2]. 강성 워터마크는 다양한 공격에도 삽입된 워터마크가 보존되어 저작물에 대한 저작권 정보를 확인할 수 있는 기술인 것에 비하여, 연성 워터마크는 작은 공격에도 워터마크가 쉽게 손상됨으로써 손상된 워터마크가 복원되지 않는 특징을 이용하여 저작물의 위조 또는 변조를 탐지하는 데 활용된다.

일반적인 오디오 워터마크 시스템은 (그림 1)과 같이 워터마크 생성, 삽입 및 검출의 3 단계로 구성된다. (그림 1)에서 점선으로 표시된 정보 흐름은 시스템 유형에 따라 선택적으로 적용되는 사항을 나타낸 것이다.

원본 워터마크는 문자열 또는 간단한 비트 열로 표현되며, 워터마크 생성 단계에서 이를 이진 비트 열로 재배열하고 추가적인 워터마크 보호를 위해 워터마크 키를 이용하여 한 번 더 처리할 수 있다. 워터마크 삽입 단계는 호스트 신호에 워터마크 신호를 삽입하는 과정으로 보안 강화를 위해 추가적인 비밀 키를 활용할 수도 있다. 이와 같이 워터마크가 삽입된 신호는 원신호와 청각적으로 유사한 품질을 제공해야 한다.

워터마크가 삽입된 신호를 전송하는 과정에서 일반적인 신호 처리 또는 악의적인 공격에 의해 워터마크의 변형이 발생할 수 있다. 수신된 신호로부터 삽입된 워터마크를 검출하는 워터마크 검출 단계에서 호스트 신호 필요 유무에 따라 non-blind 방식과 blind 방식으로 구별

된다. 강인성 측면에서는 non-blind 방식이 우수하지만, 호스트 신호를 수신 측에서 보유하기 위해서는 많은 용량의 데이터베이스가 구축되어 있어야 하는 단점으로 인해 최근에는 blind 방식의 연구가 주로 이루어지고 있다[3].

오디오 워터마크 시스템 설계 면에서 고려해야 하는 기술적 요구사항에는 비인지성, 강인성, 전송률, 복잡도 등이 있다. 특히, 강인성은 워터마크의 신뢰성을 결정하는 척도로 다양한 공격에 대한 저항력을 의미한다. 오디오 워터마크 기술에서 발생 가능한 공격의 형태로는 잡음, 리샘플링(Resampling), 재-양자화, 코덱 압축, 크로핑(Cropping), 시간 축 및 피치 변환 등이 있으며, 일부 공격은 동기화 해제로 인해 삽입된 워터마크의 생존을 심각하게 위협할 수 있다. 실제 상황에서 모든 고려사항을 만족하는 시스템의 개발은 매우 어렵기 때문에 항상 서비스 요구사항에 따른 절충점에서 타협해야 한다.

대표적인 오디오 워터마크 기술에는 LSB(Least significant bit) 방식, 반향 방식 및 대역 확산 방식 등이 있으며, 각 방식에 대해서 간략히 살펴보기로 한다.

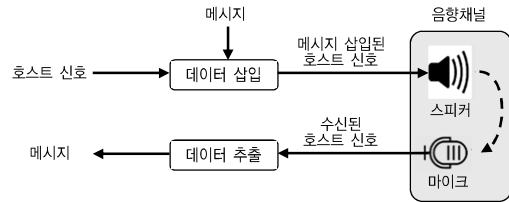
LSB 방식 오디오 워터마크는 PCM 형식의 오디오 신호에서 각 샘플의 LSB 값이 변해도 품질에는 거의 영향을 주지 않는다는 성질을 이용하여 각 샘플의 LSB를 이진수로 표시된 워터마크로 바꾸는 방식으로[4], 워터마크 삽입과 검출 과정이 매우 간단하고 품질 왜곡이 적으며 높은 전송률을 제공할 수 있다. 그러나 필터링과 코덱 압축 등의 신호처리 공격에 취약하다는 단점을 갖는다. 공격에 대한 강인성을 높이기 위해서 상위 비트에 워터마크를 삽입하거나 삽입과 검출 사이에 워터마크를 삽입하는 비트의 위치를 무작위로 정하는 방법 등이 제안되었으나[5], 신호 처리 공격에 취약한 문제점에 대한 근본적인 해결책이 되지 못하고 있다. 따라서, 주파수 영역에서 지각적으로 중요한 저주파 성분과 같은 일부 샘플의 비트에만 워터마크 변형을 가하는 방법이 사용되기도 한다.

반향 방식 오디오 워터마크는 서로 다른 오프셋과 크기를 갖는 반향 신호를 오디오 신호에 추가하는 방법으로, 구현상 간편하고 반향 신호 삽입으로 인한 잡음이 적으며 정밀한 동기화 알고리즘이 필요하지 않는 장점으로 인해 널리 사용되고 있다. 음질 왜곡을 최소화하면서 워터마크의 강인성을 높이기 위해 전-후 반향[6]과 시간-확산 반향[7] 등을 이용한 방법이 제안되었다. 반향을 이용한 워터마크 방식은 기본적으로 호스트 신호에 미치는 왜곡이 적고 여러 가지 신호 처리에 비교적 강인한 특성을 가지고 있으나 켈스트럼(Cepstrum) 연산을 통해 워터마크를 검출하기 때문에 높은 복잡도가 요구되는 것이 특징이다.

대역 확산을 이용한 오디오 워터마크 방식은 워터마크를 호스트 신호의 스펙트럼에 넓게 확산시키고 강인성을 제공할 수 있는 가장 보편적인 방식으로 알려져 있다[8]. 그러나 삽입된 워터마크가 음질 왜곡을 발생시킬 수 있으므로 심리음향 모델의 마스킹 곡선을 이용하여 워터마크의 크기를 변형한 후에 호스트 신호에 확산시키는 방법이 주로 사용되고 있다. 대역 확산 워터마크에는 직접 수열 대역 확산 방식과 주파수 호핑 대역 확산 방식이 있다. 자기 상관도는 높고 상호 상관도는 낮은 의사 난수열로 워터마크를 대역 확산하여 호스트 신호에 더하는 방법인 직접 수열 대역 확산 방식이 더 널리 사용된다. 이때, 워터마크는 시간 영역뿐만 아니라 DCT(Discrete cosine transform)를 비롯한 다양한 종류의 변환 영역에도 적용될 수 있다.

## 2. 음향 데이터 전송 기술

음향 데이터 전송 기술은 기본적으로 오디오 워터마크 기술과 유사한 기술적 배경을 가지고 출발하였다. (그림 2)는 ADT 기술의 시스템 구조를 나타낸 것으로, 오디오 워터마크 기술과 달리 송신기에 해당하는 스피커와 수신기에 해당하는 마이크로 이루어진 음향 채널



(그림 2) 음향 데이터 전송 시스템 구조

을 통해 전달된다는 점과 각종 왜곡과 공격에 대한 강인성을 제공하기 위해 정보 전달을 위한 충분한 전송률을 지원하지 못하는 오디오 워터마크에 비해 상대적으로 높은 전송률을 제공해야 한다는 면에서 차이가 있다.

ADT 기술은 메시지 정보가 삽입되는 호스트 신호의 가용 대역에 따라 비가청 대역 ADT와 가청 대역 ADT로 나눌 수 있다.

### 가. 비가청 대역 방식

비가청 대역 방식의 ADT 기술은 보통 사람이 잘 인지하지 못하는 15kHz 이상 대역 또는 그 이상의 초음파 대역에 정보 전달을 위한 메시지를 삽입하는 방식이다. 일반적인 오디오 장치들이 지원하는 48kHz 샘플링의 경우, 이론적으로 24kHz까지의 신호 대역을 표현할 수 있기 때문에 20~20,000Hz의 가청 대역을 초과하는 4kHz 대역을 데이터 전송을 위한 주파수 대역으로 활용할 수 있다. 메시지 삽입은 각 메시지 비트를 FSK(Frequency shift keying) 방식으로 변조한 후 가용 대역에 삽입하는 방식을 주로 사용한다[9].

비가청 대역 방식의 경우 비가청 대역을 사용하기 때문에 메시지 삽입으로 인한 음질 손실은 없지만, 안정적인 가용 대역 재생 및 획득을 위해서는 상대적으로 고가의 스피커와 마이크 장치가 필요하며, 24kHz 이상의 초음파 대역을 사용하는 경우에는 초음파 발생 및 수신을 위한 별도의 장비가 필요하게 되는 단점이 있다. 이러한 문제를 해결하기 위해 가용 대역을 가청 대역의 고대역까지 낮추는 방식을 사용하는 경우에는 가청 대역 방식과 같이 상당한 음질 저하를 초래하게 된다.

## 나. 가청 대역 방식

가청 대역 방식 ADT는 비가청 대역 방식과 달리 사람의 가청 대역과 메시지 삽입을 위한 가용 대역이 겹치는 경우로, 가청 대역에 메시지를 삽입함으로써 인해 상대적으로 음질 저하가 크지만, 일반적인 음향 송수신 장치로도 호스트 신호 재생 및 획득이 가능하다. 가청 대역 방식은 음질 저하에 대응하기 위한 추가적인 알고리즘이 필요하며 주로 인간의 청각 특성을 기반으로 고안된 방식을 사용한다. 대표적으로 AOFDM(Acoustic orthogonal frequency division multiplexing)과 MCLT(Modulated complex lapped transform) 방식을 비롯하여 시간 영역에서 심리음향모델을 적용한 방식과 QMF(Quadrature mirror filter) 기반 ADT 등 다양한 가청 대역 방식들이 제안되었다[10]~[15].

주파수 효율이 높은 OFDM에 기반을 둔 AOFDM 방식은 일본 NTT Docomo에서 처음 제안된 것으로, 단일 캐리어 전송 방식에 비해 다중 캐리어를 사용함으로써 평탄한 주파수 페이딩을 얻을 수 있으며, 다중 경로 하에서 프레임 간 간섭을 최소화하기 위한 GI(Guard interval)를 도입하였다. AOFDM 부호화 과정은 메시지 비트에 따라 원 신호의 고대역에 해당하는 각 캐리어 신호를 PSK(Phase shift keying)로 변조한 다음, 원 신호의 스펙트럼 포락선으로 조정하여 원 호스트 신호의 고대역을 제거한 신호에 더함으로써 AOFDM 신호를 생성한다. 수신단에서 메시지 복호화를 위해서는 AOFDM 프레임의 시작점을 식별하는 프레임 동기화 과정이 반드시 필요하므로 AOFDM 방식에서는 저대역 신호에 미리 정해진 동기화 시퀀스를 대역 확산을 통해 삽입하는 방식을 사용하였다.

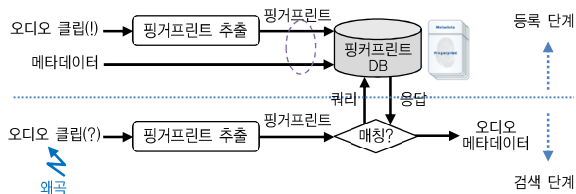
AOFDM 방식은 원 신호의 스펙트럼 변경을 통해 수십 bps 수준의 높은 전송률을 제공할 수 있지만, GI와 BPF(Band pass filter) 채용 등으로 인해 어느 정도의 왜곡은 피할 수 없는 것으로 알려져 있다.

최근 서울대를 비롯한 국내외 연구기관들에서는 MCLT를 기반으로 한 ADT 기술에 관한 연구를 활발히 진행 중에 있다[11]~[13]. MCLT는 2배로 과-샘플링된 DFT(Discrete Fourier transform)로부터 유도되므로 2배의 주파수 해상도를 제공하고, 품질 저하를 초래할 수 있는 블로킹 현상이 없는 장점을 갖는다. MCLT 방식 ADT 기술의 기본적인 전략은 MCLT 계수의 크기 성분은 변경하지 않고 위상을 수정하는 방식으로 데이터를 삽입함으로써 원음 대비 유사한 인지 품질을 얻는 데 있다. 음향 채널 왜곡에 의해 위상이 변경될 수 있지만, MCLT 계수의 실수 부는 이러한 왜곡에 영향을 받지 않는 성질로부터 수신 측에서는 실수 부의 부호로부터 메시지 복호화 과정을 수행한다. 주변 잡음에 대한 강인성을 제공하기 위해 각 메시지 비트에 대해서 대역 확장 방식을 이용하여 데이터 삽입을 수행하고, 수신된 MCLT 계수와 각 메시지 비트에 해당하는 확산 시퀀스 간의 상관도 계산을 통해 메시지 비트를 추출한다. 메시지 동기화를 위해 송신단에서는 미리 약속된 별도의 동기화 시퀀스를 삽입하고, 수신단에서는 가능한 분석 윈도우의 위치에 따른 정규화된 상관도 계산을 통해 동기화 위치를 찾는다.

## III. 음향 데이터 색인 기술 동향

인간의 지문에 착안하여 복수의 멀티미디어 객체에서 단순히 객체 자체를 비교하기보다 인지적 동일성을 설정할 수 있는 효과적인 방법으로 멀티미디어 핑거프린트가 개발되었으며, 이러한 핑거프린트는 객체의 메타데이터와 연동되어 메타데이터에 대한 인덱스로 시스템에 기여하게 된다. 핑거프린트는 객체 자체를 비교하는 방식에 비해 상대적으로 작은 크기로 인해 저장공간을 적게 필요로 하고, 인지적 무관련성이 제거되었기 때문에 효과적인 비교 및 검색이 가능하다.

일반적인 오디오 핑거프린트 시스템의 구조는



(그림 3) 오디오 핑거프린트 시스템 구조

(그림 3)과 같다.

오디오 핑거프린트 시스템은 식별 대상에 해당하는 오디오 클립으로부터 핑거프린트를 추출하고, 추출된 핑거프린트와 오디오 클립에 해당하는 메타데이터를 연계하여 핑거프린트 데이터베이스에 저장하는 등록 단계와 미지의 오디오 클립으로부터 추출된 핑거프린트를 데이터베이스에서 검색하여 일치하는 클립에 대한 메타데이터를 출력하는 검색 단계로 이루어진다.

오디오 핑거프린트 시스템을 설계하는데 고려해야 하는 파라미터에는 강인성, 신뢰성, 핑거프린트 크기, 세분성(Granularity), 검색 속도 및 확장성 등이 있다.

종래 저작권 보호에 한정되어 있던 오디오 핑거프린트 응용 분야는 2000년대 이후로 음악 및 콘텐츠 검색에 대한 요구가 증대됨에 따라 여러 가지 방식의 오디오 핑거프린트 시스템들이 개발되었으며, Facebook, Google, Microsoft, Phillips 등을 비롯한 메이저 기업뿐만 아니라 Shazam, GraceNote 등의 인터넷 음악 검색 서비스 및 솔루션 제공 업체에서도 자체적인 오디오 핑거프린트 검색 알고리즘을 보유하고 있다. 다음 절에서는 가장 대표적인 Philips 방식[16], Google waveprint[17]와 Shazam[18] 방식 오디오 핑거프린트 알고리즘에 대해서 각각 살펴보기로 한다.

### 1. Philips 방식

Philips에서 제안한 오디오 핑거프린트 방식은 비의미론적인 특징을 기반으로 검색 속도 향상을 위해 미세한 분석 구간인 프레임 단위로 해쉬 함수와 유사한 비트-스트링 형태의 부-핑거프린트를 계산하는 방식을 채

용하고 있다.

등록 및 검색 단위인 부-핑거프린트 블록은 각 프레임에서 추출된 256개의 32비트 부-핑거프린트로 이루어져 있다. 분석 프레임 간 약 97%의 큰 중첩 구간을 사용함으로써 데이터베이스의 부-핑거프린트와의 유사도 보장뿐만 아니라 뒤이은 부-핑거프린트와도 큰 유사도를 제공할 수 있다. 부-핑거프린트는 푸리에 변환, 대역 분할 및 에너지 계산, 주파수 축과 시간 축에 대한 에너지 차 신호의 부호 계산 과정 등을 거쳐 추출된다.

오디오 핑거프린트 블록을 검색하는데 걸리는 시간은 데이터베이스 규모가 커질수록 기하급수적으로 증가한다. Philips 방식에서는 검색의 효율성을 위해 256개의 부-핑거프린트로 구성된 핑거프린트 블록에서 적어도 1개 이상의 부-핑거프린트는 데이터베이스 내에 정확히 일치하는 부-핑거프린트를 갖는다는 가정하에서 개선된 검색 알고리즘을 사용하였다. 특정 부-핑거프린트와 일치하는 데이터베이스 상의 위치를 검색하기 위한 LUT(Look-up table)는 가능한 모든 32비트 부-핑거프린트 entry에 대해 데이터베이스 상의 위치 목록을 저장한 테이블로, 실제 응용에서는  $2^{32}$ 개의 entry를 갖는 LUT 대신 제한된 entry를 갖는 해시 테이블을 사용한다. 그러나, 앞서 가정이 왜곡이 심한 경우에 대해서는 실제로 유효하지 않기 때문에 추출 과정에서 얻어지는 부-핑거프린트의 비트별 신뢰도를 기반으로 왜곡의 영향을 받을 가능성에 대한 순위를 통해 검색 정확도를 개선하였다.

### 2. Google Waveprint

Google에서 제안한 waveprint 방식은 오디오 핑거프린트를 추출하는데 컴퓨터-비전에 기반한 방식을 사용하였으며, 대용량 데이터베이스에서 효과적인 이미지 쿼리를 수행하기 위해 wavelet 기반으로 접근하였다.

Waveprint 방식의 오디오 핑거프린트 추출 과정은 먼저 Philips 방식과 동일한 분석 파라미터 설정을 통해

시간 영역 오디오 신호로부터 스펙트로그램 이미지를 계산한다. 각 스펙트로그램 이미지에 대해서 wavelet을 계산한 다음, 잡음과 신호 왜곡에 대한 강인성을 제공하기 위해 크기 기반으로 정해진 개수의 wavelet들을 선택하고, 정해진 개수에 속하는 wavelet 계수와 속하지 않는 wavelet 계수를 구분하여 sparse 비트 벡터를 생성한다. Sparse 비트 벡터는 차원 감소를 위한 Min-Hash 방식을 통해 간결하고 색인 가능한 부-핑거프린트 표현으로 변환된다.

Min-Hash 방식을 통한 차원 감소에도 불구하고 다차원 공간에서 검색 과정은 상당한 계산량을 필요로 하기 때문에 효율적인 비교가 가능하고 잡음에 강인한 특성을 제공하는 LSH(Locality-sensitive hashing)를 이용한다. 일반적인 해시 과정과 달리 LSH는 부-핑거프린트의 일부만을 검사하는 일련의 해시 과정을 수행한다.

등록 단계에서 부-핑거프린트와 부-핑거프린트로부 계산된 LSH 해시 테이블을 오디오 핑거프린트 데이터베이스에 저장한다. 검색 단계에서는 등록 단계에서와 동일한 방식으로 부-핑거프린트를 계산한다. 다만 등록 단계에서는 등록 단계에서 적용한 균일한 중첩 구간 대신 무작위의 중첩 구간을 갖는다. 각 LSH 해시 테이블 성분을 이용해서 검색된 부-핑거프린트에 대한 투표를 실시하여 최소 득표수 이상을 받은 후보들을 유지하는 방식으로 후보 목록을 결정한다. 후보 목록에 포함된 모든 핑거프린트에 대해서 쿼리 부-핑거프린트와의 전체 비교를 통해 정확히 일치하는 바이트 수를 계산하여 최댓값을 갖는 부-핑거프린트를 쿼리 스펙트로그램 이미지에 대한 검색 결과로 출력한다.

### 3. Shazam 방식

Shazam 방식은 waveprint 방식과 동일하게 스펙트로그램을 기반으로 하지만, Shazam 서비스의 특성상 상당한 주변 잡음과 왜곡이 존재하는 환경에서 강인한 곡 식별이 가능하도록 스펙트로그램 피크 값을 특징으

로 사용한다. 시간-주파수 상에 위치하는 특정한 점을 중심으로 하는 해당 영역 내에서 가장 큰 에너지를 갖고 균일한 밀도를 만족하는 경우 스펙트로그램 피크로 결정된다. 원 스펙트로그램에서 피크들을 제외한 모든 점을 제거함으로써 ‘별자리 지도’라는 시간-주파수 좌표 상에 산재하는 피크 목록을 얻는다.

별자리 지도로부터 핑거프린트의 조합형 해시는 시간-주파수 쌍들의 조합으로 표현된다. 먼저 별자리 지도 상의 각 점을 앵커로 하여 각 앵커 점과 목표 지역 내 각 점과의 쌍들에 대해서 두 주파수 성분과 두 점 사이의 시차로 이루어진 해시를 생성한다. 생성된 해시는 개별 파일의 시작점에서 앵커 점까지의 시간 오프셋과 곡 식별자에 해당하는 Track ID와도 연계된다. 이러한 조합형 해시는 추가적인 저장 공간과 약간의 신호 검출 확률 손실을 희생함으로써 매우 빠른 검색 속도를 얻을 수 있다.

## IV. 음향 데이터 삽입 및 색인 서비스 동향

### 1. 음향 데이터 삽입 서비스 동향

오디오 핑거프린트와 ADT같은 음향 데이터 삽입 기술의 활용 범주는 저작권에 대한 인증 및 불법 복사 추적을 위한 저작권 보호/모니터링, 안전한 데이터 전달을 위한 비밀 키 배포, 연성 워터마크를 이용한 저작물 위변조 탐지와 각종 부가 정보 전달 수단 등이 있다.

저작권 보호와 관련해서 북미 차세대 UHD(Ultra-high definition) 방송 시스템 표준으로 채택된 ATSC 3.0은 ACR(Automatic content recognition)을 위한 워터마크 기술로 Verance의 VP1을 채택하였으며[19], 국내 ‘시상파 UHDTV송수신 정합 표준’에서는 수신기가 외부 디지털 출력을 지원하는 경우 포렌식 오디오 워터마크 기능의 추가적인 탑재를 권고하고 있다[20].

국내 IT 서비스 전문기업 카테노이드는 이러닝을 위한 동영상 플랫폼 Kollus에서 제공하는 전용 플레이어를 이용할 때, 캠코더와 같은 외부 기기를 이용하여 동



영상을 녹화하면 오디오 신호에 그 콘텐츠를 재생하는 이용자에 대한 고유 정보를 삽입하여 필요에 따라 오디오 워터마크 검출 절차를 통해 이용자를 확인할 수 있는 기능을 제공하고 있다[21].

국내 정보보안솔루션 전문 업체인 마크애니에서는 오디오 워터마크 기술을 적용한 모바일 단말 관리 솔루션을 상용화하였다. 이 솔루션을 통해 사용자는 입실 시 QR(Quick response) 코드 스캔으로 관련 앱을 설치하여 보안정책에 따라 카메라와 마이크 등의 기능을 제어할 수 있으며, 퇴실 시에는 오디오 워터마크 기술이 적용된 '사운드QR' 인식을 통해 적용된 보안정책을 해제함으로써 카메라와 마이크 등에 대한 사용을 허가할 수 있는 서비스를 출시하였다[22]. 이스라엘 공대에서는 MCLT 기반의 ADT 기술을 활용하여 휴대전화로 QR 코드를 전송하는 '음향 QR'을 제안하기도 하였다[13].

NFC(Near-field communication) 방식과 유사하게 Zoosh, Paytm, Ultracash, ToneTag, Chirp 등 음향 신호를 기반으로 한 다양한 근거리 지불 솔루션들이 개발되었으나[23], 대부분 아직 널리 활용되지는 못하고 있는 실정이다. 미국 모바일 커머스 업체인 Shopkick은 오프라인 가맹점에서 방문 고객을 대상으로 상품 정보, 쿠폰 및 포인트를 제공하기 위한 수단으로 바코드 방식과 함께 ADT 기술을 채용하였으며[24], 국내 SK플래닛에서 Shopkick을 자회사로 인수하여 자사의 시럽 서비스에 적용하려는 계획을 발표한 사례가 있다.

국내 스타트업 기업인 사운드리는 자사의 비가청 음파를 통한 정보 전달 기술을 활용하여 극장 내 영화 예고편과 스마트폰 연계 서비스, 커피 매장에서 주문 가능한 사운드 비콘, TV 광고와 연계한 다양한 마케팅과 광고 효과 측정 서비스 등을 개발하였다[25].

## 2. 음향 데이터 색인 서비스 동향

오디오 핑거프린트 기술의 활용 범주는 미지의 오디오에 대한 식별, 오디오 데이터의 변경 여부를 확인하는

무결성 검증, 추출된 핑거프린트를 워터마크의 비밀키로 활용하는 워터마크 지원, 핑거프린트로부터 서로 다른 수준의 정보를 추출하는 검색 등이 있다.

오디오 핑거프린트를 활용한 음악 검색 서비스로 Shazam, SoundHound, Gracenote MusicID와 ACR Cloud 등이 있다. 가장 대표적인 Shazam 서비스는 오디오 핑거프린트를 기반으로 음악과 TV 프로그램뿐만 아니라 광고까지 검색할 수 있으며, 이를 통해 음원/프로그램 구매를 유도하고 SNS 공유 기능, 관련 정보 제공 및 추천 기능 등을 지원하고 있다. 국내 매장 음악 서비스 사업자인 원트리즈뮤직은 2016년 8월에 Shazam과 음악 검색 관련 빅데이터 수급 계약을 통해 K-Pop 수출 역량 강화를 도모하고 매장 음악 선곡 절차에 Shazam 서비스를 도입할 예정이라고 밝혔으며[26], 삼성전자는 2017년 CES에서 자사의 스마트TV에 Shazam을 탑재하여 방송 시청 중에 흘러나오는 음악을 찾아주는 서비스를 선보였다[27].

시청률 조사 서비스와 관련해서 최근 시청 매체의 다양화와 함께 종래의 시청률 조사 방식에서 다루기 어려운 PC 및 스마트폰을 통한 TV 시청과 VoD 시청 등 새로운 시청 행태에 대한 조사를 위해 닐슨과 TNmS 등 시청률 전문 조사기관에서 오디오 핑거프린트를 채용한 시청률 조사 방식을 개발 중에 있다.

한편, MPEG-7 오디오 표준은 오디오 신호의 본질적 특징들을 요약한 오디오 데이터의 색인을 위한 포괄적인 프레임워크를 제안하여 오디오에 대한 고유한 식별자인 핑거프린트를 추출할 수 있는 기능을 제공하며, 장시간 데이터 교환을 위해 오디오를 기술하고 색인하는 메타데이터 구조를 XML(eXtensible markup language) 기반으로 제안하였다[28].

## V. 맺음말

본고에서는 콘텐츠에 대한 접근 편의성 향상을 위한



음향 신호 기반 데이터 삽입 및 색인 기술 연구 그리고 관련 기술을 활용한 서비스 동향에 대해서 살펴보았다.

콘텐츠 관련 시장이 최근 급격하게 성장함에 따라 콘텐츠 관리 및 제어/식별 분야에 대한 기술적 요구사항 또한 활발하게 이루어지고 있으며, 이러한 추세를 반영하고 ADT와 같은 기술적 진보를 기반으로 종래 콘텐츠 서비스의 개선뿐만 아니라 향후 다양한 분야에서 해당 기술을 접목한 신규 서비스를 발굴할 수 있는 계기가 될 것으로 예상된다.

#### 용어해설

**오디오 워터마크** 저작권 보호를 위해 사람의 귀에 의해서 인지되지 않도록 음성, 음악 등 오디오 신호에 정보를 삽입하는 기술

**오디오 핑거프린트** 음원에 포함된 고유한 일부 특징을 추출하여 데이터베이스에 있는 원본의 특징과 비교하여 연결된 메타데이터 정보를 제공하는 기술

**음향 데이터 통신** 음향 신호의 가청 또는 비가청 주파수 대역에 특정 코드나 데이터를 삽입함으로써 스피커와 마이크를 통한 정보 전달 수단을 제공하는 기술

#### 약어 정리

ATSC	Advanced Television Systems Committee
ACR	Automatic Content Recognition
ADT	Acoustic Data Transmission
AOFDM	Acoustic Orthogonal Frequency Division Multiplexing
BFP	Band Pass Filter
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DRM	Digital Rights Management
FSK	Frequency Shift Keying
GI	Guard Interval
LSB	Least Significant Bit
LSH	Locality-Sensitive Hashing
LUT	Look-Up Table
MCLT	Modulated Complex Lapped Transform
NFC	Near Field Communication
PCM	Pulse Code Modulation
PSK	Phase Shift Keying

QMF	Quadrature Mirror Filter
QR	Quick Response
UHD	Ultra High Definition
XML	eXtensible Markup Language

#### 참고문헌

- [1] F. Hartung and M. Kutter, "Multimedia Watermarking Techniques," *Proc. IEEE*, vol. 87, no. 7, July 1999, pp. 1079-1107.
- [2] 오병택, 문병주, 이동일, "디지털 워터마킹 기술동향 및 전망," *전자통신동향분석*, 제17권 제6호, 2002. 12, pp. 155-162.
- [3] Y. Lin and H.A. Waleed, "Audio Watermarking Techniques," in *Audio Watermark*, Switzerland: Springer International Publishing, 2015, pp. 51-94.
- [4] W. Bender et al., "Techniques for Data Hiding," *IBM Syst. J.*, vol. 35, no. 3-4, 1996, pp. 313-336.
- [5] W. Cao, Y. Yan, and S. Li, "Bit Replacement Audio Watermarking using Stereo Signals," *Int. Conf. New Trends Inform. Service Sci.*, Beijing, China, June 30-July 2, 2009, pp. 603-606.
- [6] H.J. Kim and Y.H. Choi, "A Novel Echo-hiding Scheme with Backward and Forward Kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, Aug. 2003, pp. 885-889.
- [7] B.-S. Ko, R. Nishimura, and Y. Suzuki, "Time-Spread Echo Method for Digital Audio Watermarking," *IEEE Trans. Multimedia*, vol. 7, no. 2, Apr. 2005, pp. 212-221.
- [8] M. Arnold, M. Schmucker, and S.D. Wolthusen, *Techniques and Applications of Digital Watermarking and Content Protection*, Norwood, MA, USA: Artech House, 2002.
- [9] P. Bihler, P. Imhoff, and A.B. Cremers, "SmartGuide-A Smartphone Museum Guide with Ultrasound Control," *Procedia Comput. Sci.*, vol. 5, 2011, pp. 586-592.
- [10] H. Matsuoka et al., "Acoustic OFDM: Embedding High Bit-Rate Data in Audio," *International Conference on Multimedia Modeling, Lecture Notes in Computer Science*, Heidelberg, Berlin: Springer, 2008.
- [11] H.S. Yun, K. Cho, and N.S. Kim, "Acoustic Data Transmission Based on Modulated Complex Lapped Transform," *IEEE Signal Process. Lett.*, vol. 17, no. 1, 2010, pp. 67-70.
- [12] R. Frigg et al., "Acoustic Data Transmission to Collaborating Smartphones—an Experimental Study,"

- Annu. Conf. Wireless On-demand Netw. Syst. Services*, Obergurgl, Austria, Apr. 2-4, 2014, pp. 17-24.
- [13] I. Dagan, G. Binyamin, and A. Eilam, "Delivery of QR Codes to Cellular Phones Through Data Embedding in Audio," *IEEE Int. Conf. Sci. Electr. Eng.*, Eilat, Israel, Nov. 16-18, 2016, pp. 1-4.
- [14] S. Beack, T. Park, and K. Kang, "Acoustic Data Transmission Based on Complex Quadrature Mirror Filterbank Transform," *IEEE Int. Symp. Consumer Electron.*, Jeju, Rep. of Korea, June 22-25, 2014, pp. 1-2.
- [15] S. Beack et al., "Acoustic Data Transmission by Extension on the Time Domain Approach," *IEEE Global Conf. Signal Inform. Process.*, Orlando, FL, USA, Dec. 14-16, 2015, pp. 1111-1115.
- [16] J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System," *Int. Conf. Music Inform. Retrieval*, Paris, France, Oct. 13-17, 2002.
- [17] S. Baluja and M. Covell, "Waveprint: Efficient Wavelet-Based Audio Fingerprinting," *Pattern Recog.*, vol. 41, no. 11, 2008, pp. 3467-3480.
- [18] A. Wang, "An Industrial Strength Audio Search Algorithm," *Int. Conf. Music Inform. Retrieval*, Baltimore, ML, USA, Oct. 26-30, 2003.
- [19] A/334:2016, "ATSC Standard: Audio Watermark Emission (A/334)," ATSC, Sept. 2016.
- [20] TTAKKO-07.0127, "지상파 UHDTV 방송 송수신 정합," TTA표준화 위원회, 2016. 6.
- [21] 디지털뉴스부, "카테고이드, 클라우드형 온라인 비디오 플랫폼 (OVP) 출시," 디지털 타임스, 2013. 3. 19.
- [22] 성지은, "마크애니, 사생활 침해없는 MDM 솔루션 공개," 아이뉴스, 2016. 8. 30.
- [23] E. Mesropyan, "Sound-based Payments as an Inclusive Technology for the Developing World," Oct. 28, 2016. <https://letstalkpayments.com/sound-based-payments-as-an-inclusive-technology-for-the-developing-world/>
- [24] Shopkick, Accessed 2017. <http://www.shopkick.com/how-it-works>
- [25] Soundly's Official Blog, Accessed 2017. <http://blog.soundly/>
- [26] 연합뉴스, "원트리즈뮤직, 음악 검색 앱 사장과 빅데이터 수급 계약," 연합뉴스, 2016. 8. 24.
- [27] 강일용, "[김조한의 미디어 세상] 삼성전자, 마침내 스마트TV를 정의하다," IT동아, 2017. 1. 31.
- [28] O. Hellmuth et al., "Using MPEG-7 Audio Fingerprinting in Real-world Applications," *Audio Eng. Soc. Convention*, New York, USA, Oct. 10-13, 2003, pp. 1-10.