

논문 2013-50-12-26

오디오 핑거프린팅기반 입체음향 재현 시스템

(Audio Fingerprinting Based Spatial Audio Reproduction System)

류 상 현*, 김 형 국**

(Sang Hyeon Ryu and Hyoung-Gook Kim[©])

요 약

본 논문에서는 오디오 핑거프린팅 방식과 스페셜 오디오 처리 방식을 결합한 오디오 핑거프린팅 기반 입체음향 재현 시스템을 제안한다. 제안된 시스템에서는 변조스펙트럼 기반의 명확한 오디오 정점 핑거프린트를 이용하여 잡음환경에서 오디오 핑거프린팅 시스템의 검색정확도를 향상시켰으며, 메타데이터로 제공되는 스페셜 오디오 정보는 청취자에게 소리가 실제로 녹음된 공간에서 소리를 듣는 것 같은 느낌을 준다.

Abstract

This paper proposes a spatial audio reproduction system based on audio fingerprinting that combines the audio fingerprinting and the spatial audio processing. In the proposed system, a salient audio peak pair fingerprint based on modulation spectrum improves the accuracy of the audio fingerprinting system in real noisy environments and spatial audio information as metadata gives a listener a sensation of being listening to the sound in the space, where the sound is actually recorded.

Keywords : Audio Fingerprinting, Modulation Spectrum, Spatial Audio, Metadata

I. 서 론

인간의 삶과 밀접한 관계를 유지하면서 인간의 사상과 감정을 표현하는 음악은 IT 기술의 발달과 함께 새로운 서비스의 형태로 진화하고 있다.

특히, 다양한 서비스 중에서 음악 식별을 통한 검색

기술과 보다 나은 음질을 제공하기 위한 입체음향 재현 기술에 대한 관심도가 증가되고 있다.

음악 식별 기술은 음악의 신호특징을 이용하여 임의의 음악을 찾아내는 기술로서 오디오 핑거프린팅 기술^[1]로 알려져 있으며 방송 음악 모니터링^[2]을 자동으로 수행하거나 모바일 기기 등을 통해 어딘가에서 들려오는 음악신호에 대한 음악 정보 제공 서비스^[3] 또는 P2P 등에서 불법 음악 공유를 방지하기 위한 저작권 보호 기술^[4]로서도 활용되고 있다. 오디오 핑거프린팅 기술은 무엇보다도 다양한 실제 잡음환경에서 녹취된 짧은 길이의 오디오 신호로부터 추출된 오디오 핑거프린트를 이용하여 빠르고 정확도가 높은 검색 성능을 제공해야 한다. 이러한 검색 성능을 제공하기 위해 개발된 오디오 핑거프린팅 방식을 보유하고 있는 기업으로는 Gracenote, Shazam, Google 등이 대표적이다.

* 학생회원, ** 정회원 광운대학교 전자공학과
(Department of Wireless Communication
Engineering, Kwangwoon University)

© Corresponding Author(E-mail: hkim@kw.ac.kr)

※ 본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음(NIPA-2013- H0301-13-4005).

※ 이 논문은 2013년도 정보(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업지원을 받아 수행된 것임(NRF-2013R1A1A2007601).

접수일자: 2013년10월22일, 수정완료일: 2013년11월27일

입체음향 재현기술^[5]은 사용자에게 음악의 원음장을 충실하게 재현하는 것뿐만 아니라 음악의 음원이 발생한 공간에 위치하지 않은 청취자가 음향을 들었을 때 방향감, 거리감 및 공간감을 느낄 수 있도록 공간정보를 부가하는 기술로서 5.1 채널 이상의 채널수를 가지는 3차원 오디오 기술로 진화되고 있으며, MPEG-H ISO/IEC 23008^[6]로 현재 표준화가 진행 중이다. 특히, 코덱자체보다는 다양한 출력채널 환경과 스피커의 배치 환경에 적응적인 렌더링이 3차원 오디오 표준에서 중요한 부분을 차지할 것으로 예측된다. 또한, MPEG Surround^[7]나 Spatial Audio Object Coding^[8]에서 표준화된 방식들이 중요한 역할을 하리라 기대된다. 이와 함께, 최근에는 입체 비디오와 같이 현장감과 실재감 그리고 높은 몰입도를 제공하기 위한 스페셜 오디오 서비스(Spatial Audio Service)를 적용한 기술 개발이 진행 중이다.

본 논문에서는 지금까지 각각 독립적으로 적용 및 개발되고 있는 음악 식별 기술과 입체음향 재현기술을 결합하여 새로운 서비스를 제공할 수 있는 오디오 핑거프린팅 기반 입체음향 재현 시스템을 제안하고자 한다.

제안된 시스템을 통하여 사용자는 어디선가 들려오는 음악의 가수, 곡명 등의 음악정보와 이미 이 음악을 들은 다른 사용자들의 평가의견들도 제공받을 수 있을 뿐만 아니라, 검색된 음악을 듣기를 원할 때 사용자가 보유하고 있는 디바이스에 맞도록 입체음향을 제공받을 수가 있다.

본 논문의 구성은 다음과 같다. II장에서는 오디오 핑거프린팅 기술과 스페셜 오디오 재현 기술을 결합한 오디오 핑거프린팅 기반 입체음향 재현 시스템에 대해서 기술하고, 이와 함께 제안된 시스템의 오디오 핑거프린팅 기술에 있어서 현재 개발된 Gracenote, Shazam, Google 등의 오디오 핑거프린팅 기술을 분석하고, III장에서는 대표적인 세 가지 방식을 기반으로 최근에 개발된 다양한 오디오 핑거프린팅 기술들에 대한 검색 성능을 제안된 방식과 비교 평가하였다. 그리고 IV장에서는 결론 및 향후 연구 방향을 제시하였다.

II. 본 론

1. 오디오 핑거프린팅을 이용한 입체음향 재현 시스템의 전체 구조

본 연구에서는 그림 1에서와 같이 오디오 핑거프린팅 기술과 스페셜 오디오 재현기술을 결합한 오디오 핑거프린팅 기반 입체음향 재현 시스템을 제안한다.

오디오 핑거프린팅 기반 입체음향 재현 시스템의 전체 구조는 사용자의 단말기와 서버 영역으로 크게 구분된다.

먼저, 사용자의 단말기에서는 음원소스로부터 몇 초간의 음악부분구간을 녹취하고, 녹취된 음악신호의 신호특징을 이용하여 오디오 핑거프린트를 추출하는 오디오 핑거프린트 생성부(Audio Fingerprint Generation), 생성된 오디오 핑거프린트를 압축하여 서버로 전송한 후에 서버로부터 전송받은 음원 정보(음악의 곡명, 사용자 리뷰 등), 그리고 음원 데이터와 입체음향 재현 메타데이터를 이용하여 사용자의 단말기에 맞게 최대한 입체음향을 재생하는 오디오 재생부(Audio Player)로 구성된다. 그리고 서버에서는 각 음원 데이터를 저장한 음원 데이터베이스(Audio Source Database), 곡명, 가수명 등의 각 음원의 정보와 함께, 인터넷 커뮤니티에 작성된 음원을 청취한 사용자의 청취 후기 등의 정보를 포함한 오디오 정보 데이터베이스(Audio Information Database), 음원 데이터베이스로부터 각 음원 데이터 전체의 오디오 핑거프린트를 추출하여 저장한 오디오 핑거프린트 데이터베이스(Audio Fingerprint Database), 오디오 핑거프린트 데이터베이스와 단말기로부터 전송받은 단 구간 오디오 쿼리의 핑거프린트를 매칭하여 음원을 검색하는 오디오 핑거프린트 매칭부(Audio

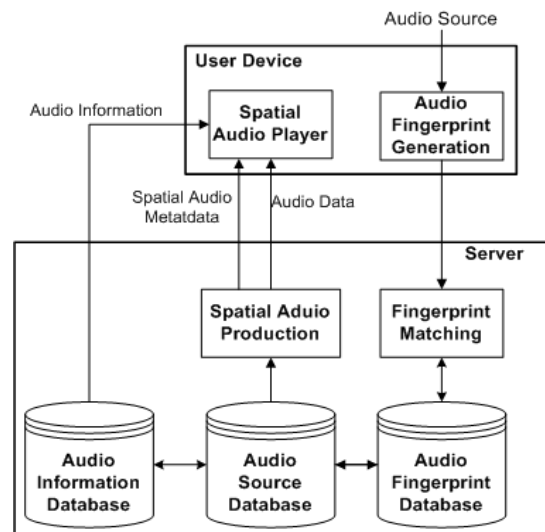


그림 1. 전체 시스템 구조
 Fig. 1. System architecture.

Fingerprint Matching), 음원 데이터로부터 녹음환경에 대한 정보가 담겨져 있는 메타데이터를 생성하는 스페셜 오디오 제작부(Spatial Audio Production)로 이루어져 있다.

본 연구에서 제안하는 시스템은 다음과 같은 과정을 거쳐 수행된다. 먼저, 사용자가 임의의 음원 쿼리 일부분을 몇 초간 자신의 단말기로 녹음한다. 녹음된 음원 쿼리의 신호특징을 이용하여 오디오 핑거프린트를 추출하고 오디오 쿼리 핑거프린트로 사용한다. 추출된 오디오 쿼리의 핑거프린트는 서버로 전송되어 서버의 음원 핑거프린트 데이터베이스에 저장되어 있는 오디오 핑거프린트와 매칭되어 사용자의 오디오 쿼리에 대한 음원 검색 결과 및 사용자의 음원에 대한 리뷰정보를 제공받는다. 이와 함께, 사용자가 검색된 음원을 듣고자 하면 음원 검색 결과에 해당하는 음원의 데이터를 음원 데이터베이스로부터 찾고, 스페셜 오디오 제작부에서 형성된 해당 음원의 입체음향 메타데이터를 압축된 음원과 함께 사용자의 단말기로 전송한다. 사용자 단말기의 오디오 재생부에서는 전송받은 압축된 음원의 데이터를 복원하고, 복원된 음원과 입체음향 메타데이터를 이용하여 현재 사용자 단말기의 출력장치에 적합한 형태로 오디오 신호를 재구성하여 입체음향을 출력시킨다. 또한, 음원을 재생하는 동안 서버로부터 전달받은 음원의 정보를 사용자 단말기의 디스플레이 장치에 표시하여 사용자가 가시적으로 음원의 정보를 읽을 수 있도록 한다.

2. 오디오 핑거프린트 추출 방식의 비교

다양한 잡음환경에서 높은 검색성능을 획득하기 위해 개발된 대표적인 오디오 핑거프린팅 기술은 Gracenote, Shazam, Google의 방식이다.

Philips 혹은 Gracenote의 오디오 핑거프린팅 방식^[4]은, 오디오 신호의 스펙트럼을 로그스케일의 33개의 서브밴드로 분할하고, 분할된 각 서브밴드의 에너지 차이를 이진화하여 오디오 핑거프린트를 생성한다. 그리고 서브 밴드로 중첩된 오디오 핑거프린트를 보상하기 위해서 쿼리에서 찾아진 원래의 서브밴드 핑거프린트의 1 bit 에러의 해밍거리만큼 해시값을 확장하여 검색하기 때문에, 잡음환경에서의 오디오 왜곡에 강인한 성능을 제공한다.

Google의 오디오 핑거프린팅 방식^[9]은, 오디오 신호

의 웨이브렛에 의한 변환 주파수값을 이미지화하여 에너지의 피크값(정점값)을 양자화하는 오디오 핑거프린트를 추출한다. 그리고 유사한 특징을 가지는 핑거프린트를 효율적으로 저장하는 해싱방법을 사용하기 때문에, 메모리를 효율적으로 사용하고 잡음환경에 강인한 성능을 제시한다.

Shazam의 오디오 핑거프린팅 방식^[10]은, 오디오 신호의 Short Time Fourier Transform(STFT)을 통한 주파수 스펙트럼에서 추출된 정점간의 위치정보인 Landmark(랜드마크)를 이용하여 오디오 핑거프린트로 추출하기 때문에 잡음환경에 강인한 성능을 보인다. 또한, 정점간의 위치정보를 이용한 해싱방법을 사용함으로써 효율적이고 빠른 검색 성능을 제공한다.

하지만, 위 세 가지 방식들은 모두 오디오 쿼리의 녹취 길이가 긴 경우에는 검색에서 있어서 좋은 성능을 보이지만, 녹취 길이가 짧으면 오디오 쿼리를 식별할 충분한 정보가 없기 때문에, 잡음 및 에코 환경이 비교적 높은 환경에서는 오디오 신호의 왜곡으로 인하여 검색 성능이 저하되는 문제점이 발생한다. Philips의 방식에서는 서브밴드에 대한 특징정보의 양이 현저하게 줄어들기 때문에 잡음에 강인한 에너지 정점(피크값)을 추출하기가 어렵다. 또한 Google의 방식에서도 에너지의 정점을 양자화기 때문에 특징정보의 양이 줄어들어 잡음환경에 강인하지 못한 성능을 보인다. 그리고 Shazam의 방식에서는 짧은 오디오 쿼리에서 잡음으로 인한 주파수 스펙트럼의 왜곡으로 인해 정점 추출이 어려워져 검색 성능의 저하가 발생한다.

본 논문에서는 이러한 문제를 해결하고, 짧은 오디오 쿼리에서 잡음 및 에코 환경에 따른 오디오 신호의 왜곡에 강인한 오디오 핑거프린트를 추출하기 위해서 변조 스펙트럼^[11]을 적용한다. 즉, Shazam의 방식을 기반으로 변조 스펙트럼으로부터 정점의 위치정보를 추출하여 오디오 핑거프린트를 추출하는 방식을 제안한다. 사용자 단말기에서는 녹음된 부분구간 음원의 오디오 신호에서 변조 스펙트럼에서의 정점 위치정보를 기반으로 오디오 핑거프린트를 생성하고, 생성된 오디오 핑거프린트를 압축하여 서버로 전송한다.

다음 그림 2는 오디오 프린트를 생성하는 과정을 나타내는 블록도이다.

먼저, 오디오 핑거프린트를 추출하기 위해 입력된 오디오 신호의 변조 스펙트럼(Modulation Spectrum) 계

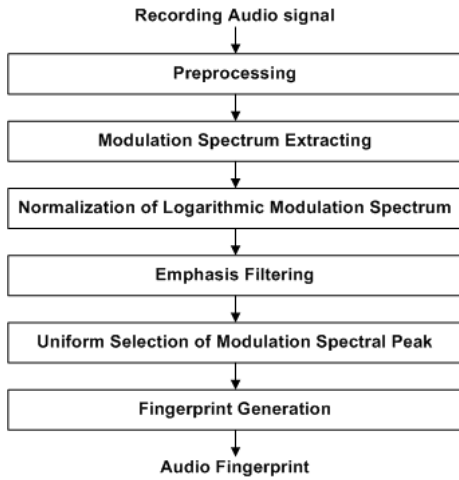


그림 2. 오디오 핑거프린트 추출 구조도
 Fig. 2. Block diagram of audio fingerprint extraction.

수를 생성한다. 생성된 변조 스펙트럼 계수에 로그스케일을 적용하고, 평균을 차감하는 방식을 통해서 로그스케일의 변조 스펙트럼을 정규화 한다. 그리고 감쇄 왜곡에 민감한 고주파의 스펙트럼 성분을 강화하기 위해서 정규화된 로그 스케일의 변조 스펙트럼에 강화 필터를 적용한다. 강화 필터링된 변조 스펙트럼에서 대해 역동적인 문턱값이 적용된 순방향 및 역방향 필터링을 적용함으로써 잡음환경에 보다 강인한 정점들을 검출한다. 검출된 정점 중에서 가장 인접한 정점 쌍을 핑거프린트 혹은 랜드마크라고 정의한다. 랜드마크는 한 쌍의 피크점에 대한 주파수의 시작점과 끝점, 두 피크점간의 시간차이 그리고 랜드마크의 시작시간을 핑거프린트 정보로 저장한다. 생성된 오디오 핑거프린트 정보는 20 bit의 해시 값으로 변환되고 서버로 전송되어 서버의 오디오 핑거프린트 데이터베이스의 핑거프린트와 매칭 과정을 거쳐 음원의 ID를 획득한다.

3. Spatial Audio기반의 입체음향 제작 방식

Spatial Audio기반의 입체음향 제작의 목적은 사용자의 오디오 쿼리에 대해 서버에서의 매칭 과정을 거쳐 검색된 음원의 ID에 해당하는 음원의 입체음향 메타데이터를 이용하여 사용자가 실재감 있는 오디오 서비스를 제공받도록 하기 위함이다.

그림 3은 서버에서 수행되는 스파셜 오디오 제작과정의 블록도이다.

음원 소스 데이터베이스로부터 각 음원 데이터가 입력되면, 인코더에서는 메타데이터 엔진으로부터 입력받

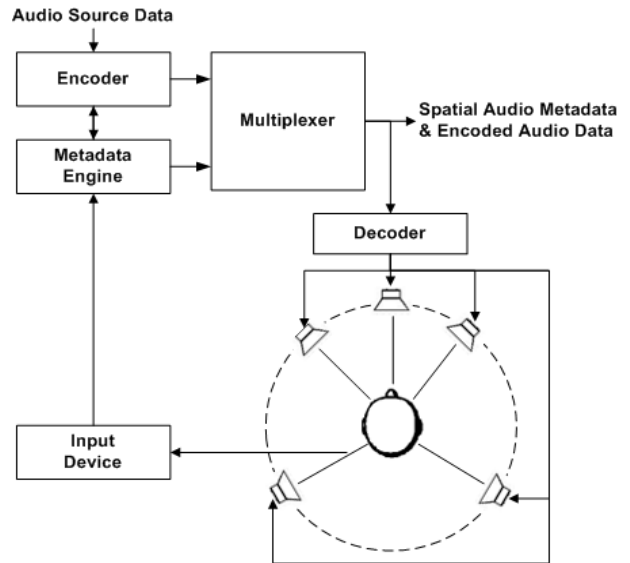


그림 3. 스파셜 오디오 프로덕션 구조도
 Fig. 3. Block diagram of Spatial audio production.

은 입력장치에 대한 음원의 메타데이터를 통해 인코딩한다. 메타데이터 엔진에서는 입력장치에 따라서 음원 데이터를 콘트롤하여 사용자에게 실재감을 전달할 수 있는 입체음향 생성을 위해 스파셜 오디오 방식을 적용하여 채널신호의 수, 스피커의 위치에 따라 녹음되는 가장 큰 에너지를 갖는 객체의 에너지 값, 객체들 간의 에너지 비, 객체들 간의 유사도 등의 입체음향 파라미터 정보를 입체음향 메타데이터로 생성한다. 생성된 메타데이터는 인코더와 멀티플렉서로 전달된다. 인코딩된 오디오 데이터와 메타데이터는 멀티플렉서를 통해 결합된 데이터 형태로 합쳐지고, 결합된 데이터는 사용자의 단말기로 전송된다. 동시에, 결합된 데이터는 서버의 모니터링 디코더로 전송된다. 생성된 스파셜 오디오 데이터와 메타데이터를 가상의 출력 시스템으로 구동하여 실재감 있는 오디오 효과가 제대로 구현되었는지를 판단한다. 모니터링 디코더를 통해서 얻어진 오디오 효과를 바탕으로 실시간으로 메타데이터 엔진에 적용하여 더 정확한 메타데이터를 생성하기 위해서 엔진을 제어한다. 이 피드백 과정을 통해 보다 실재감 있는 오디오 신호가 사용자의 단말기로 전송된다.

4. 스파셜 오디오기반의 입체음향 재생

사용자 단말기의 오디오 재생기에서는, 서버로부터 전송받은 오디오 데이터와 스파셜 오디오 메타데이터를 통해서 사용자의 단말기 환경에 적합한 형태로 오디오

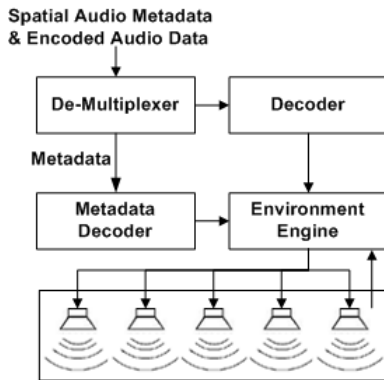


그림 4. 스페셜 오디오 재생 구조도
Fig. 4. Block diagram of spatial audio playing.

데이터를 재구성한다. 사용자의 단말기마다 출력환경이 다르기 때문에, 사용자의 단말에 적합한 스페셜 오디오를 구성해야한다.

그림 4는 사용자 단말기의 오디오 재생기에서 수행되는 스페셜 오디오 재생 과정의 블록도이다.

먼저, 서버로부터 전송된 오디오 데이터와 메타데이터는 디멀티플렉서를 통해서 분리된다. 분리된 오디오 데이터는 디코더를 통해서 오디오 신호로 디코딩되고, 메타데이터는 메타데이터 디코더를 통해서 디코딩 된다. 디코딩된 오디오 신호와 메타데이터는 환경엔진을 통해서 사용자의 출력 환경에 따라서 실시간으로 재구성된다. 재구성된 스페셜 오디오 신호는 사용자 단말기의 출력 스피커 시스템을 통해 사용자에게 실제감 있는 오디오 서비스를 제공한다.

III. 실험 결과

본 논문에서 제안한 시스템의 성능을 평가하기 위해서 먼저, 오디오 핑거프린팅의 검색 성능을 측정하였다. 또한, 제안된 오디오 핑거프린팅(MOD)의 성능을 측정하기 위해서 최근에 개발된 네 개의 오디오 핑거프린팅 방식과 성능을 비교하였다. 방식 1(MW)은 STFT기반 정점 추출을 통한 오디오 핑거프린팅 방식^[10], 방식 2(MS)는 피치 추출을 기반으로 서브핑거프린트의 마스크를 사용하는 오디오 핑거프린팅 방식^[12], 방식 3(MC)은 마스크된 오디오 스펙트럼의 키포인트를 기반으로 오디오핑거프린트를 추출하는 방식^[13], 그리고 방식 4(MX)는 적응적으로 스케일된 오디오 신호의 시간-크로마 표현방식으로부터 추출된 로컬 특징값 추출 방식

^[14]이다. 적용된 방식 1, 방식 2, 방식 3, 방식 4는 모두 논문의 내용을 기반으로 구현하여 실험에 적용하였다.

실험을 위해 사용된 두 가지 테스트 데이터베이스 유형은 다음과 같다 : (1) 데이터베이스 1: 팝, 힙합, 재즈, 클래식 등 다양한 장르의 7000곡으로 구성했고, (2) 데이터베이스 2: 10~15분 길이의 740시간 길이의 4000개 TV 광고로 구성했다.

모든 오디오 데이터는 휴대폰과 같은 휴대용 단말을 고려한 실제 오디오 데이터로부터 16 kHz 샘플링레이트로 변환된 모노, 16 bit로 저장되어 오디오 핑거프린트 추출에 적용되었다. 오디오 쿼리는 TV에 연결된 2.1채널 스피커로부터 5m 떨어진 지점에서 휴대용 단말로 녹음된 2, 3, 4, 5초 길이의 오디오 데이터를 사용했다. 다양한 잡음환경에서의 성능측정을 위해 무작위로 생성된 3000개 오디오 쿼리의 음악신호에 Babble, Moving Car, White, Street, Computer Pan 등의 5 종류의 잡음을 clean 쿼리에서부터 0 dB, 6 dB, 12 dB의 신호 대 잡음비(SNR)로 구성하여 잡음환경에 대한 오디오 핑거프린트 검색 성능을 측정했다. 오디오 쿼리는 16 kHz 샘플링, 16bit로 양자화된 모노채널인 표준 PCM 형태로 구성되었다.

표 1은 데이터베이스 1에서 5초길이의 쿼리를 사용하여 제안된 방식과 최근에 개발된 네 가지 방식에 대한 실험결과를 비교한 결과표로서, 다섯 가지 잡음환경에서의 결과에 대한 평균이다.

표 1에 의하면 제안된 방식이 clean 및 잡음환경에서 비교된 네 가지 방식보다 모두 우수한 성능을 보임을 알 수 있다.

표 2는 데이터베이스 2에서 수행된 실험결과이다. 광고 음원에 대한 인식 성능이 음악검색에 대한 인식 성능 보다는 약간 저하된 이유는 광고 내에 묵음구간이 존재하기 때문이다. 그러나 광고음원에 대한 인식성능 역시 제안된 방식이 비교된 네 가지 방식들의 인식 성

표 1. 데이터베이스 1에 대한 실험결과 비교
Table 1. Comparison of experimental results with database 1.

SNR	Averaged Recognition Rate (%)				
	MOD	MW	MS	MC	MX
clean	97.3	95.5	96.7	94.8	93.5
12dB	96.8	93.7	95.4	89.6	78.9
6dB	93.6	88.4	85.6	77.5	63.8
0dB	80.7	73.6	71.9	61.7	57.6
Total	92.1	87.8	87.4	80.9	73.5

표 2. 데이터베이스 2에 대한 실험결과 비교
Table 2. Comparison of experimental results with database 2.

SNR	Averaged Recognition Rate (%)				
	MOD	MW	MS	MC	MX
clean	95.5	93.6	94.5	93.5	92.6
12dB	94.3	90.5	92.8	86.2	75.4
6dB	91.6	85.4	81.7	74.6	60.2
0dB	77.5	70.8	67.9	58.5	53.4
Toal	89.7	85.1	84.2	78.2	70.4

표 3. 제안방식의 쿼리 길이에 따른 실험결과
Table 3. Experimental results of the proposed method according to query length.

SNR	Averaged Recognition Rate (%)			
	2초	3초	4초	5초
clean	76.8	91.5	95.1	97.3
12dB	71.5	90.7	94.3	96.8
6dB	63.3	84.7	91.8	93.6
0dB	55.6	76.5	81.7	80.7
Toal	66.8	85.9	90.7	92.1

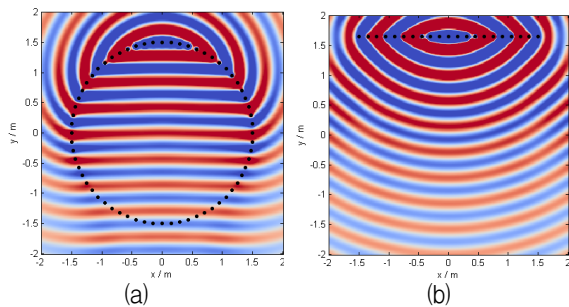


그림 5. 스파셜 오디오 시뮬레이션 결과
Fig. 5. Simulation results of spatial audio processing.

능 보다 좋은 결과를 제시하고 있음을 알 수 있다.

표 3은 제안된 방식의 쿼리 길이에 따른 인식률 결과를 나타낸다. 표 3에 나타난 바와 같이 오디오 쿼리 길이가 길어지면 성능이 향상된다는 것을 알 수 있으며, 제안된 방식이 4, 5초 길이의 쿼리 및 SNR 6dB이상의 실험환경에서 모두 90% 이상의 우수한 인식률을 보임을 알 수 있다.

그림 5는 녹음환경에서 획득된 메타데이터와 출력스피커의 형태에 따른 입체 오디오 효과를 시뮬레이션한 결과이다. 모든 오디오 데이터는 휴대폰과 같은 휴대용 단말을 고려한 실제 오디오 데이터로부터 16 kHz 샘플링, 16 bit로 양자화된 PCM형태로 구성되었다. 그림 5(a)는 60개의 스피커를 원형으로 배치했을 때의 입체 오디오 효과를 시뮬레이션한 결과다. 원형의 스피커 공

간 안에서 음향이 고루 전달되며, 실재감 있는 음향 효과를 나타내고 있다. 그림 5(b)는 앞선 시뮬레이션과는 다른 환경인 20개의 스피커를 일렬로 배치했을 때의 입체 오디오 효과를 시뮬레이션한 결과다. 일렬로 배치된 스피커에서 출력된 음성 신호가 전방위로 퍼져나가는 형태를 보이고 있다. 스피커의 전방에서 사용자가 입체 음향 효과를 체험 할 수 있음을 보여준다.

두 시뮬레이션 결과그림은 출력스피커의 형태에 따라서 디코딩된 음원에 메타데이터를 반영함으로써 각 디바이스의 구조에 맞는 실재감 있는 음향효과를 보이고 있다.

IV. 결론 및 향후 연구

본 논문에서는 오디오 핑거프린팅 기술과 입체음향 기술을 결합한 오디오 핑거프린팅을 이용한 입체음향 재현 시스템을 제안했다. 제안된 시스템은 오디오 핑거프린팅을 이용하여 녹음된 음원을 식별하고, 식별된 음원의 녹음환경에 대한 입체음향 메타데이터와 음원을 사용자에게 전송하고, 사용자의 단말에서는 전송받은 메타데이터를 사용하여 사용자의 출력환경에 적합한 형태의 스파셜 오디오를 재구성하여 사용자가 실재감 있는 오디오를 청취할 수 있도록 하였다. 제안된 시스템의 오디오 핑거프린트 방식은 기존의 방식들에 비해 우수한 성능을 보이며, 실제 휴대용 단말에 적용되어 사용되기 적합함을 보여준다. 또한, 출력 스피커의 형태에 따라 음장감 있는 음향효과를 보여줌으로써, 제안된 시스템은 향후, 실제 휴대용 단말기에 입체음향 어플리케이션의 형태로 적용되어 서비스되기에 적합한 시스템이라고 판단된다.

REFERENCES

- [1] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of algorithms for audio fingerprinting," IEEE Workshop on Multimedia Signal Processing, pp. 169-173, Dec. 2002.
- [2] J. Cerquides, "A real time audio fingerprinting system for advertisement tracking and reporting in FM radio," Radioelektronika, 2007. 17th International Conference, pp. 1-4, Apr. 2007.
- [3] A. Sinitsyn, "Duplicate song detection using

audio fingerprinting for consumer electronics devices,” IEEE International Symposium on Consumer Electronics, pp.1-6. Jun. 2006.

- [4] J. Haitzma, and T. Kalker, “A highly robust audio fingerprinting system,” 3rd International Conference Music Information Retrieval, pp. 107-115, Oct. 2002.
- [5] J. I. Seo, and G. O. Gang, “Introduction and Standard Status of High Order Multichannel AudioSystem for Realistic Audio Broadcasting”, Electronics and Telecommunications Trends, pp. 49-56, Dec. 2012.
- [6] ISO/IEC 23008, Information technology, “High efficiency coding and media delivery in heterogeneous environments,” MPEG-H, Jun. 2013.
- [7] I. S. Jang, S. K. Beack, J. I. Seo, and D. Y. Jang, “Overview of MPEG Surround,” Journal of Broadcast Engineering, pp. 181-190, Jun. 2006.
- [8] J. I. Seo, I. S. Jang, and G. O. Gang, “Overview of object based audio technology in MPEG,” Korea Society Broadcast Engineers Magazine, pp. 75-87, Jun. 2009.
- [9] S. Baluja, and M. Covell, “Beyond “near-duplicates”: learning hash codes for efficient similar-image retrieval,” 20th International Conference on Pattern Recognition, Aug. 2010.
- [10] A. Wang, “An industrial strength audio search algorithm,” 4th International Conference Music Information Retrieval, pp. 7-13, Oct. 2003.
- [11] H. -G. Kim, K. W. Um, J. Y. Kim, Y. Y. She, and X. Zhu, “Method, Medium, and system for music retrieval using modulation spectrum,” US Patent, US 7818167 B2, Oct. 2010.
- [12] X. Pan, X. Yu, J. Deng, W. Yang, and H. Wang, “Audio fingerprinting based on local energy centroid,” IET International Communication Conference on Wireless Mobile and Computing, pp. 351-354, Nov. 2011.
- [13] X. Anguera, A. Garzon, and T. Adamek, “MASK: robust local feature for audio fingerprinting,” International Conference on Multimedia and Expo, pp. 455-460, Jul. 2012.
- [14] M. Malekesmaeli, and R. K. Ward, “A novel local audio fingerprinting algorithm,” IEEE 14th International Workshop on Multimedia Signal Processing, pp. 136-140, Sep. 2012.

— 저 자 소 개 —



류 상 현(학생회원)
2012년 광운대학교 전파공학과
공학사
2012년~현재 광운대학교
전파공학과 석사과정
<주관심분야 : 음성 및 오디오 신
호처리, 햅틱 신호처리, 멀티미디
어 신호처리>



김 형 국(정회원)
1999년~2002년 독일 SIEMENS
/ Cortologic AG 책임연구
원
2002년~2005년 독일 베를린
공과대학교 Assistant
Professor
2005년~2007년 삼성종합기술원 수석연구원
2007년~현재 광운대학교 전파공학과 교수
<주관심분야 : 음성 및 오디오 신호처리, 햅틱 신
호처리, 멀티미디어 신호처리>