한국음향학회지 제28권 제4호 pp. 375~382 (2009)

압축 도메인 특징을 이용한 강인한 오디오 핑거프린팅

Robust Audio Fingerprinting Using Compressed-Domain Features

서 진 수*, 이 승 재**
(Jin Soo Seo*, Seungjae Lee**)

*강릉원주대학교 전자공학과, **한국전자통신연구원 SW 콘텐츠 연구부문
(접수일자: 2009년 3월 11일; 채택일자: 2009년 4월 18일)

본 논문에서는 압축도메인 특징을 이용한 오디오 핑거프린팅 방법을 제안하였다. 압축도메인을 이용함으로써 계산량과 시간을 크게 줄일 수 있는 장점이 있다. 특히 오디오 압축에 널리 쓰이고 있는 MDCT 도메인을 이용하였으며, MDCT 도메인을 부밴드로 나누고 대표적인 모멘트 특징인 에너지, 무게중심, 평탄도로 부터 각각 핑거프린트를 얻었다. 추출된 특징을 차분 필터링하고 부호를 취하여 이진 핑거프린트를 얻었다. 실험을 통해서 고려한 MDCT 도메인 특징들로부터 얻은 핑거프린트들의 인식 성능을 비교하였다. 수 천곡 규모의 오디오에 대해서 다양한 변환에 대한 인식 성능을 고려하였으며, 실험 결과 부밴드 에너지가 가장 우수한 핑거프린팅 성능을 보였다.
핵심용어: 오디오 핑거프린팅, 오디오 인식, 부밴드 에너지, 부밴드 무게중심, 부밴드 평탄도
투고분야: 음향 신호처리 분야 (1.2)

This paper proposes a new audio fingerprinting method based on compressed-domain features. By basing on the compressed domain, the computational efficiency of the proposed method can be greatly enhanced. Especially we deal with MDCT domain, which is widely employed in audio compression, and extract three kinds of subband features; energy, centroid, and flatness. By taking signs after differentially filtering each feature, binary audio fingerprints are obtained. The identificaton performance of the three kinds of fingerprints are experimentally compared. Among the considered compressed-domain subband features, the subband energy showed the best performance for fingerprinting.
Keywords: Audio Fingerprinting, Audio Identification, Subband Energy, Subband Centroid, Subband Flatness
ASK subject classification: Acoustic Signal Processing (1.2)

I. 서 론

정보 처리 기기 및 기술의 발달에 따라 콘텐츠 산업은 사용자의 편의를 증대시키기 위해서 양방향화, 휴대화, 지능화 되는 방향으로 발전하고 있다. 따라서 사용자의 요구를 능동적으로 반영하여, 오디오를 제공하는 것을 가능하게 하는 오디오 정보 처리 및 검색 기술의 중요성이 커가고 있다 [1-3]. 본 논문에서는 오디오 정보 검색 기술 중 오디오 인식에 대해서 다루며, 특히 압축된 오디오 파일에서 고속으로 특징을 추출하는 방법을 제안한다. 오디오 인식은 오디오 핑거프린팅 (fingerprinting) 또는 해싱 (hashing) 이라고도 불리며, 오디오의 고유한 특징을 이용하여 해당 오디오를 인식한다. 이 때 사용되는 특징을 핑거프린트라 한다. 일반적으로 오디오 인식기에 사용되는 핑거프린트는 다음의 4요소를 만족시켜야 한다 [4].

- 차별성 (pairwise independence): 서로 다른 음악에 대해서 오인식이 일어나지 않도록 충분한 차별성을 가지고 있어야함
- 강인성 (invariance to distortions): 오디오 신호가 압축, EQ, 잡음첨가, sampling rate 변화 등 다양한 변환을 겪어 신호에 변화가 가해지더라도 그 값이 일정한 범위 내에서 유지되어야함
- 간결성 (compactness): 다수의 오디오에서 핑거프린트를 추출해서 저장하므로, 작은 크기의 표현이 필요함

책임저자: 서 진 수 (jsseo@kangnung.ac.kr)
210-702 강원도 강릉시 강릉대학로 120 강릉원주대학교 전자공학과
(전화: 033-640-2428; 팩스: 033-646-0740)

• 계산용이성 (computational efficiency): 핑거프린트 추출에 있어서 계산량과 걸리는 시간이 작아야함

대부분의 디지털 오디오 데이터는 압축 파일 형태로 저장 및 판매되고 있지만, 기존의 오디오 특징 추출 방법들은 그림 1 (a)와 같이 복호화한 후 다시 오디오를 분석하여 특징을 추출한다. 일반적으로 분석 과정은 복호화 및 전처리, FFT 등의 주파수 변환, 특징 추출, 후처리 등으로 이루어진다. 계산량 또는 시간이 제약되는 휴대용 단말 (MP3P, PMP, 뮤직폰 등), AOD (Audio on Demand) 등의 환경에서 오디오 인식을 적용하기 위해서는 복호화 후에 다시 오디오를 분석하는 것은 비효율적인 부분이 있다. 오디오 압축은 오디오를 분석하여 복원에 중요한 정보만을 남겨 놓은 효율적인 표현 방식이다. 따라서 오디오 압축도메인에서도 충분히 정보를 추출하여 인식 등의 오디오 정보처리를 할 수 있다. 비디오의 경우 압축도메인을 이용한 정보처리 연구는 상대적으로 많이 연구되어왔다 [5]. 오디오 압축도메인을 이용하여 계산용이성을 증대시키기 위해서, 본 논문에서는 그림 1 (b)와 같이 압축 파일을 완전히 복호화 하지 않고 압축 도메인 상에서 다른 변환 등의 추가적인 신호 분석 과정 없이 직접 핑거프린트를 추출하는 방법을 제안한다. 특히 가장 널리 사용되고 있는 오디오 압축 방법인 MP3 (MPEG-1 Layer III)에 대해 알아보고, MP3 파일로부터 오디오를 완전히 복호화 하지 않고 압축도메인에서 핑거프린트를 추출하는 방법을 제안한다. 본 논문에서는 여러가지 MP3 파일 요소들 중에서 신호의 고유 특징을 가장 잘 나타내는 MDCT (modified discrete cosine transform) 계수를 이용한다. MDCT 계수는 MP3 뿐 아니라 MPEG-2 AAC 등 다양한 오디오 압축 방법의 근간이 되므로 [6] 본 논문에서 제시된 방법은 약간의 파라미터 수정 후에 다른 압축 파일들에도 적용할 수 있다. 본 논문에서는 MDCT 특징으로 부밴드 에너지 (subband energy, SE) [7], 부밴드 무게중심 (subband centroid, SC) [8], 부밴드 평탄도 (subband flatness, SF) [9]의 세 가지를 고려하였다. 얻어진 특징을 시간-주파수 차분 필터를 통해 가공하여 이진 핑거프린트를 얻었다 [7]. 실험을 통해서 본 논문에서 제안한 압축 도메인 오디오 핑거프린트는 계산용이성이 향상되고, 간결성, 차별성, 강인성을 만족할 수 있음을 보였다.

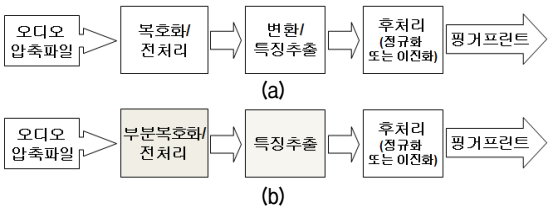

그림 1. 오디오 핑거프린트 추출 과정 블록선도
(a) 기존 방법, (b) 압축도메인 방법

Fig. 1. Block diagram of fingerprint extraction.
(a) Conventional approach, (b) Compressed-domain approach

본 논문에서는 MP3 압축 도메인 특징인 MDCT로부터 오디오 인식을 위한 특징인 핑거프린트를 얻는 방법을 제안하였다. II장에서 MP3 압축과 MDCT 변환에 대해서 간략히 살펴보고, III 장에서 제안된 압축 도메인 핑거프린트 추출 방법을 기술하고, IV장에서 제안된 핑거프린트들의 오디오 인식 성능을 실험하고 결과를 비교 분석한다.

II. MP3 압축과 MDCT 변환

오디오 압축도메인 특징을 이용하여 핑거프린트를 추출하기 위해서는 오디오 부호화와 복호화에 대한 이해가 필요하다. 본 장에서는 가장 널리 사용되고 있는 오디오 압축 방법인 MP3와 다양한 압축 표준에서 널리 쓰이고 있는 MDCT 변환에 대해서 살펴본다.

2.1. MP3 부호화 및 복호화 개요

오디오를 MP3 압축 방식으로 부호화하는 방법은 그림 2와 같이 요약할 수 있다. 먼저 PCM오디오 신호를 32개의 같은 간격의 부밴드로 나누어주는 필터뱅크 (filter-bank)를 거치고 각각의 부밴드 신호에 대해서 18차수의 MDCT를 수행한다 [6]. 이렇게 두 번에 걸쳐서 변환을 수행하는 이유는 MDCT 변환을 사용하지 않고 필터뱅크만을 이용한 기존 MPEG-1 Layer I, II 오디오와 호환성을 가지기 위해서이다. MPEG-1과의 호환성이 필요 없는 MPEG-2 AAC (Advanced Audio Coding) 부호화의 경우에는 필터뱅크과정 없이 바로 MDCT를 수행한다. 32개의 부밴드에서 18개의 계수들이 나오므로 하나의 프레임은 총 576개의 MDCT 계수로 이루어진다. 이 MDCT계수들에 대해서 사람의 청각 모델을 적용하여 가장 청각에 민감한 대역에 많은 비트를 할당한다. 이러한 비트 할당은 호프만 코드 테이블, global gain, scalefactors (각 critical 밴드 대역에서의 noise shaping factor) 등을 정해 주어서, 허용된 비트레이트 (bit rate) 내에서 양자화 오차가 청각 모델에서 얻은 매스킹 문턱값 (masking threshold) 보다 작도록 만들어 준다. 이 과정은 보통 합성에 의한

분석 방법 (analysis-by-synthesis)을 통해서 반복적으로 비교해 가면서 최적의 호프만 코드 테이블, global gain, scalefactors 값 등을 찾게 된다. 청각 모델과 비트 할당 방법은 MPEG 표준안으로 강제되지 않으며 최종 비트 스트림 포맷이 표준안과 일치하면 된다. 따라서 MP3 부호화기마다 음질 차이가 나게 되고, 이는 사용된 청각 모델과 비트 할당 방법이 다르기 때문이다. MPEG-1 Layer 3 표준에서 허용하는 비트레이트에는 32, 40, 48, 56, 64, 80, 96, 112, 128, 144, 160, 192, 224, 256, 320 kbit/s가 있고, 허용하는 샘플링 주파수 (sampling rate)에는 32, 44.1, 48 kHz가 있다. 여기에 MPEG-2 와 MPEG-2.5 표준으로 추가된 MP3 포맷의 비트레이트에는 8, 16, 24, 144 kbit/s가 있고 역시 추가된 샘플링 주파수에는 8, 11.025, 12, 16, 22.05, 24 kHz 가 있다.

MP3 복호화 과정은 그림 3에 나온 바와 같이 부호화 과정의 역으로 진행된다. 먼저 MP3 헤더의 syncword를 찾아서 비트 스트림의 동기를 찾은 후에 호프만 코드 테이블, scalefactor 값 등 MP3 복호화에 필요한 데이터들을 찾는다. 호프만 디코딩을 통해 MDCT 값을 얻고 이를 양자화 과정에서 곱한 scalefactor 값으로 나누는 descaling 과정을 거친다. 최종적으로 IMDCT (Inverse MDCT)와 역 필터뱅크 (inverse filterbank) 과정을 통해서 오디오 신호를 복원해 내게 된다.

2.2. MDCT 변환과 스펙트럼 분석

MDCT 변환은 MP3 압축뿐만 아니라 MPEG-2 AAC 등 다양한 오디오 코덱에 사용되고 있는 변환 방식이다. MDCT는 lapped transform의 일종으로 프레임마다 50%씩 overlap이 되어있고 $2M$ 개의 데이터에 대한 MDCT 결과는 M 개의 계수로 주어진다. 반대로 IMDCT 과정에서는 $2M$개의 계수에서 M개의 데이터를 만들어 내는데 연속된 프레임에서의 IMDCT값을 이용해야 원래의 오디오 값을 복원하는 것이 가능하다. 즉 최소한 연속한 2개의 프레임의 MDCT 계수가 있어야 원래의 신호를 복원가능한 것이 lapped transform의 특징이다. 수식으로 살펴보면 MDCT 변환은 입력 신호 $x[n]$과 MDCT 베이시스 $hk[n]$ 대해서 식 (1)과 같이 주어지고, IMDCT 변환은 현재와 이전 프레임의 MDCT 값인 $X[k]$와 $X^{P}[k]$로부터 식 (2)와 같이 주어진다 [6].

$$X[k] = \sum_{n=0}^{2M-1} x[n] h_k[n] \tag{1}$$

$$x[n] = \sum_{k=0}^{M-1} \left(X[k] h_k[n] + X^{P}[k] h_k[n+M] \right) \tag{2}$$

이러한 MDCT와 IMDCT의 관계를 Y. Wang et al. [11]의 연구 결과를 바탕으로 고찰해 보면 신호에 MDCT를

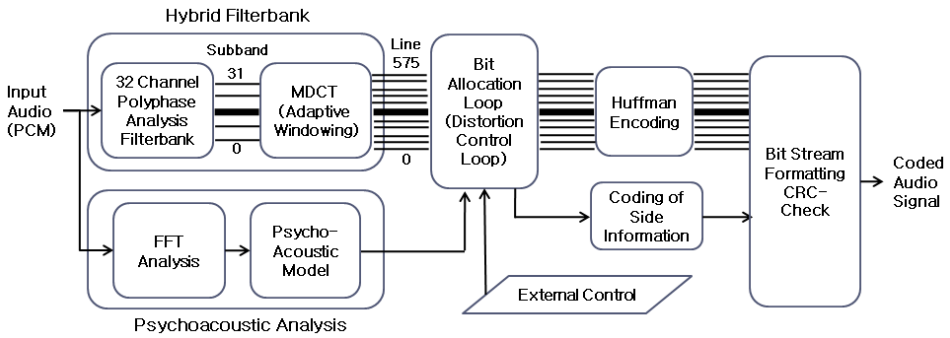

그림 2. MP3 오디오 압축 부호화 과정 블록선도 [10]
Fig. 2. Block diagram of MP3 audio compression [10].

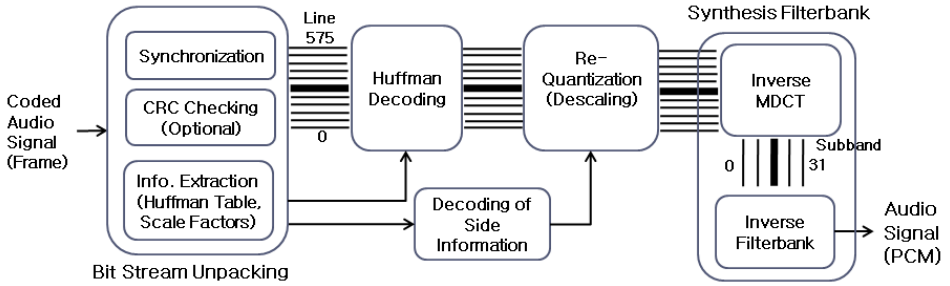

그림 3. MP3 오디오 압축 복호화 과정 블록선도 [10]
Fig. 3. Block diagram of MP3 audio decompression [10].

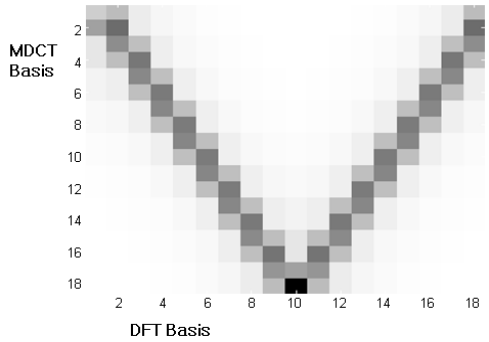

그림 4. MDCT 베이시스의 주파수 특성
Fig. 4. Frequecny response of MDCT basis.

가하는 것은 DFT (Discrete Fourier Transform)를 시간과 주파수 방향으로 각각 u, v 만큼 이동하는 SDFT$_{u,v}$ (Shifted DFT)을 가하는 것과 같다. 실수의 신호에 대해서는 식 (3)과 같이 신호 $x(n)$의 MDCT값인 $X(k)$는 시간 방향으로 $(M+1)/2$, 주파수 방향으로 1/2만큼 이동한 SDFT의 실수부와 같다.

$$X[k] = Real\left\{SDFT_{\frac{M+1}{2},\frac{1}{2}}(x[n])\right\} \tag{3}$$

식 (3)을 보면 신호의 SDFT 계수 중 실수부의 정보만이 MDCT에 남아 있게 된다. 따라서 MDCT값만으로는 DFT처럼 완전한 오디오 주파수 특성 분석에 사용하기에는 부족하지만, MDCT 변환의 베이시스가 정현파 형태로서 그림 4에서 볼 수 있듯이, 각각의 베이시스를 DFT 변환을 취해서 살펴보면 주파수 상에 국부적으로 강하게 응집되어 있어 주파수 특성 분석에 사용 가능함을 알 수 있다.

III. 압축도메인 특징 기반 오디오 핑거프린팅

3.1. 압축 도메인 핑거프린트 추출

본 장에서는 MP3 압축도메인의 MDCT 값으로부터 직접 특징을 추출하여 핑거프린트를 구하는 과정에 대해서 살펴본다. 그림 1(a)에서 살펴본 바와 같이 일반적으로 오디오 핑거프린팅에서는 전처리 과정에서 샘플링 주파수를 미리 정해준 특정값으로 맞추어 주고 특징 추출을 위한 오디오 신호 분석을 하게 된다 [4]. 본 논문에서 목표로 하고 있는 MP3 압축도메인의 경우에는 부호화 시에 샘플링 주파수와 무관하게 일정한 길이의 MDCT 변환을 취한다. 따라서 MP3 파일의 샘플링 주파수 차이를 보완해 주지 않으면, 같은 음악이라 하더라도 서로 다른 샘플링 주파수로 MP3 압축을 할 경우 크게 다른 MDCT 값을 가지게 된다. 따라서 그림 5에 주어진 바와 같이 MDCT 값을 바로 이용하는 것이 아니라 시간 방향으로 샘플링 주파수에 대해서 정규화를 먼저 수행한다. 이때 입력 오디오가 2채널 스테레오일 경우 두 채널의 MDCT 값을 합하여 모노로 만든 후 샘플링 레이트에 대한 정규화를 수행한다. 시간 방향 정규화 과정은 인접한 W_{Ref} 개의 MP3 프레임을 하나의 윈도우로 보고 W_{Ref} 개의 MDCT값의 제곱의 평균을 구해서 (FFT 크기의 제곱으로 스펙트럼을 구하는 것과 같이 제곱을 취함) 그 윈도우의 MDCT 파워스펙트럼으로 사용한다. 다음 윈도우에 대해서도 마찬가지로 윈도우 내의 MDCT 제곱의 시간방향 평균으로 MDCT 파워스펙트럼을 구한다. 각 샘플링 주파수 별로 가변 윈도우의 길이는 표 1과 같다. MP3 압축에서는 샘플링 주파수에 무관하게 항상 고정된 길이 (하나의 프레임에 576 샘플)의 MDCT 변환을 취하므로 시간과 주파수 윈도우의 길이가 일부 샘플링 주파수에서는 정수값을 가지지 못한다. 이를 정수값으로 반올림할 경우 오디오 데이터에 시간축변환 (speed change)을 [12] 가한 것과 같은 효과를 가지게 되어서 성능 면에서 큰 저하가 있다. 따라서 본 논문에서는 표준 시간 윈도우 길이 W_{Ref} 를 실수로 하고, n 번째 시간 윈도우의 길이 $W(n)$을 식 (4)에 주어진 바와 같이 반올림 함수 $\partial\{\}$을 이용하여 가변함으로써 성능의 저하를 줄였다.

$$W(n) = \partial\{(n+1)\times W_{Ref}\} - \partial\{n\times W_{Ref}\} \tag{4}$$

시간방향 정규화 수행 후에 주파수 방향으로 MDCT 파워스펙트럼 계수 들을 표 1과 같은 주파수 경계를 가진 부밴드들로 나누고, 각 부밴드 내에서 특징을 구한다. 이

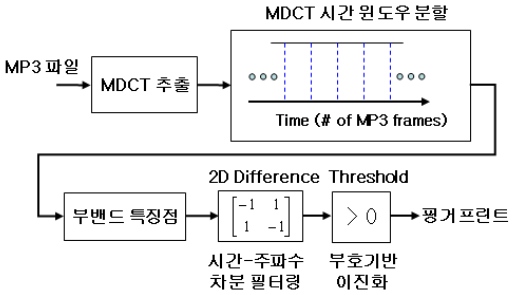

그림 5. 제안된 MP3 압축도메인 MDCT 계수 기반 오디오 핑거프린트 추출 블록선도
Fig. 5. Block diagram of the proposed fingerprint extraction method based on the MP3 MDCT coefficients.

표 1. MP3 MDCT 계수 표준 시간 윈도우 길이 W_{Ref} 및 부밴드 주파수 경계
Table 1. Temporal window length and subband frequency boundary for MP3 MDCT.

샘플링 주파수	W_{Ref}	부밴드 주파수 경계 $S[m]$, $0 \le m \le 17$ (MDCT 계수 위치)
11025	2	{0, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96, 104, 112, 120, 128, 136}
16000	2.9025	{0, 6, 11, 17, 22, 28, 33, 39, 44, 50, 55, 61, 66, 72, 77, 83, 88, 94}
22050	4	{0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68}
32000	5.8050	{0, 3, 6, 8, 11, 14, 17, 19, 22, 25, 28, 30, 33, 36, 39, 41, 44, 47}
44100	8	{0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34}
48000	8.7075	{0, 2, 4, 6, 7, 9, 11, 13, 15, 17, 18, 20, 22, 24, 26, 28, 29, 31}

때 n번째 시간 윈도우의 k번째 MDCT 값의 제곱의 평균인 $P[n, k]$에 대해서 m번째 부밴드의 에너지 $E[n, m]$, 무게중심 $C[n, m]$, 평탄도 $F[n, m]$ 특징은 부밴드 주파수 경계 $S[m]$에 대해서 각각 식 (5), (6), (7) 과 같이 주어진다.

$$E[n,m] = \sum_{k=S[m]+1}^{S[m+1]} P[n,k] \tag{5}$$

$$C[n,m] = \frac{\sum_{k=S[m]+1}^{S[m+1]} kP[n,k]}{E[n,m]} \tag{6}$$

$$F[n,m] = \frac{\left[\prod_{k=S[m]+1}^{S[m+1]} P[n,k]\right]^{1/N_m}}{E[n,m]/N_m} \tag{7}$$

위 (7) 에서 $N_m = S[m+1]-S[m]$ 으로 주어진다. 이어서 얻어진 각각의 윈도우의 부밴드 특징을 그림 5에 주어진 바와 같이 인접한 시간-주파수 상에서 차분 필터를 [7] 통과시키고 그 결과의 부호값을 취하여 핑거프린트 이진화를 수행하였다.

3.2. 특징 비교

본 연구의 주목적은 MP3 압축도메인의 MDCT 부밴드 특징의 오디오 인식 성능을 검증하는 것이다. 특징 비교 방법은 성능 비교를 용이하게 하기위해서 3.1절에서 설명한 바와 같이 기존의 논문 [7]과 동일하게 특징을 이진화하고 해밍 거리를 사용하였다. 일반적으로 오디오 인식 문제는 오디오 특징 추출 함수인 H와 특징간 거리 비교 함수인 D를 이용하여 아래와 같은 가설검증 (hypothesis testing)으로 주어진다 [4].

가설 L0: 만약 D(H(A), H(B)) 이 문턱값 T보다 작다면, 두 오디오 클립 A와 B는 같은 오디오이다.
가설 L1: 만약 D(H(A), H(B)) 이 문턱값 T보다 크다면, 두 오디오 클립 A와 B는 다른 오디오이다.

각각의 MP3 MDCT 시간 윈도우를 표 1에 나온 바와 같이 17개의 부밴드로 나누어 시간-주파수 차분 필터링 후에 윈도우 마다 16비트의 핑거프린트가 얻어진다. 하지만 1개의 윈도우에서 추출된 16비트의 핑거프린트로는 다양한 오디오 변형들에 대해서 강인성을 유지하면서 위 가설검증을 엄밀하게 적용하는 것이 불가능하다. 따라서 인접한 N개의 윈도우에서 나온 특징들을 모아서 16xN 비트를 모아서 핑거프린트 비교에 사용하게 된다 [4,8]. 본 논문에서는 약 10초 길이의 MP3 파일의 핑거프린트를 인식에 사용하여 N = 100 이며, 윈도우당 16비트가 추출되므로 비교에는 총 1600 비트가 사용된다. 특징 비교를 위해 해밍 거리를 아래와 같이 두 오디오 특징 블록 p와 q 간의 거리 비교에 사용하였다.

IV. 실험 결과 및 고찰

본 장에서는 MP3 압축의 MDCT 도메인 부밴드 핑거프린트의 오디오 인식 성능을 비교한다. 실험을 위해서 수천곡 분량의 다양한 장르의 MP3 음악 파일을 수집하였다. 수집된 MP3 파일들에서 III장에서 고려한 세가지 MDCT 도메인 부밴드 특징들을 각각 이용하여 핑거프린트 DB를 만들어 성능 실험을 수행하였다.

오디오 핑거프린트는 I장에서 기술한 바와 같이 차별성, 강인성, 간결성, 계산용의성의 네 가지 성질을 만족해야 한다. 각각의 성질에 대해서 실험 또는 분석을 통해서 본 논문에서 고려한 부밴드 에너지, 무게중심, 평탄도 핑거프린트들의 성능을 비교하겠다. 일반적으로 인식 문제에는 두 가지 형태의 오인식율이 있다. FAR (false alarm rate)은 서로 다른 오디오를 같다고 판정할 확률이며, FRR (false rejection rate)는 같은 오디오를 다르다고 판정할 확률이다 [4]. 두 가지 오인식율이 모두 작아야 하며, 핑거프린트의 차별성은 FAR과 강인성은 FRR과 연관되어 있다. 일반적으로 두 오디오 간의 특징 비교를 위한 가설검증에 사용되는 문턱값 T를 작게하면 FAR은 작아지고 FRR은 커지고, 반대로 T를 크게하면 FAR은 커지고

FRR은 작아진다. 따라서 일반적으로 원하는 FAR값을 만족시키는 T값을 정하고, 그 때의 FRR값을 구하여 성능비교를 수행한다.

4.1. 간결성과 계산용이성 고찰

간결성의 경우 제안된 방법의 핑거프린트는 이진화 되어 있고, 표 1에 제시한 특징 윈도우 길이를 사용할 경우 1초당 평균 9.57개의 윈도우에서 16차의 이진 특징 벡터가 얻어지므로 153.13 bps (bits per second) 정도의 핑거프린트 값으로 입력 MP3 파일의 정보를 요약할 수 있다. 보통의 오디오 압축에서 수십 kbps 이상의 데이터가 필요하므로 얻어진 MP3 핑거프린트는 아주 간결하다고 할 수 있다.

계산용의성의 경우 그림 1에 제시한 바와 같이 제안된 방법은 압축 도메인을 사용하므로, 기존의 핑거프린팅 방법들에서 압축된 음악 파일을 복호화 하고 전처리 후에 신호 특성을 분석하기 위해서 퓨리에 변환 등을 취해야 하는 등의 부가적인 계산 과정들을 생략할 수 있어 계산량을 줄일 수 있다. 제안된 방법은 그림 3의 MP3 복호화 과정 중에서 Inverse MDCT와 Inverse Filterbank 과정을 생략할 수 있고, 그림 1 (a)의 핑거프린트 추출 과정 중 변환 부분을 생략할 수 있다.

4.2. 차별성과 강인성 성능 검증

서로 다른 오디오에서 얻어진 핑거프린트들 간의 거리 비교를 통해서 차별성을 확인해 보도록 하겠다. 실험은 임의로 선택된 400,000쌍의 서로 다른 10초 길이의 오디오 세그먼트들간의 핑거프린트 거리 비교를 통해 이루어졌다. 그림 6은 400,000 개의 해밍 거리 값들의 히스토그램이다. 특징별로 살펴보면 부밴드 에너지와 무게중심은 평균값인 0.5를 중심으로 표준편차가 각각 0.0137과 0.0144로 조밀하게 분포하고, 반면에 부밴드 평탄도의 경우 평균값인 0.497을 중심으로 표준편차가 0.0166로 상대적으로 성기게 분포하고 있다. 그림 6을 보면 해밍 거리 값을 실험에서 얻은 평균과 표준편차로부터 구한 정규분포와 비교해보면 근사적으로 잘 맞음을 알 수 있다 [7, 8]. 이로부터 핑거프린트 간의 해밍 거리 D를 정규분포로 가정하면, D의 평균 m_D과 표준편차 σ_D에 따라서 FAR은 식 (8)과 같이 주어진다.

$$FAR = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi\sigma_D}} exp\left[-0.5\left(\frac{x-m_D}{\sigma_D}\right)^2\right] dx \quad (8)$$
$$= 0.5\, erfc\left(\frac{m_D - T}{\sqrt{2}\,\sigma_D}\right)$$

표 2에 식 (8)을 이용하여 다양한 FAR 값에 대해서 역으로 특징 비교 문턱값인 T를 구하였다.

핑거프린트의 강인성에 대한 실험적인 검증을 위해서 DB 상의 원본 오디오 데이터를 낮은 레이트의 MP3 압축 (32 kbps), 3 dB EQ (이퀄라이제이션), 선형속도변환 (LSC), time-scale modification (TSM) 등의 다양한 변형을 가하였다. 낮은 레이트의 MP3 압축을 제외한 다른 변형들은 변형을 가한 후에 다시 MP3 압축 (128 kbps)을 하였다. 각각의 훼손된 버전의 MP3 파일에서 같은 방식으로 핑거

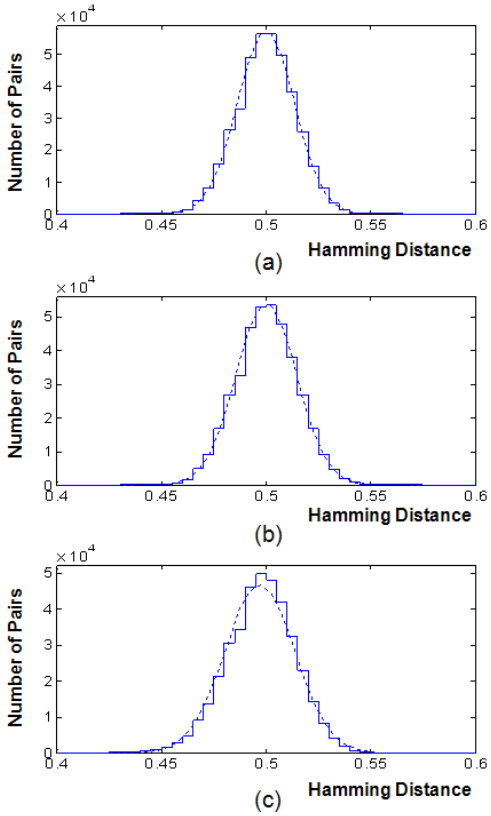

그림 6. 서로 다른 오디오의 핑거프린트들 간의 해밍 거리 히스토그램; 실선: 실험결과, 점선: 정규분포모델링
(a) 부밴드 에너지, (b) 부밴드 무게중심, (c) 부밴드 평탄도

Fig. 6. Histogram of the Hamming distance between the fingerprints from different audio blocks; solid line: experimental result, dashed line: modeling with normal distribution.
(a) subband energy, (b) subband centroid, (c) subband flatness

표 2. FAR 값에 해당하는 핑거프린트 검증 문턱값 T
Table 2. Threshold T for recognition corresponding to FAR values.

특징\FAR	10^{-20}	10^{-18}	10^{-16}	10^{-14}	10^{-12}
에너지	0.3734	0.3803	0.3876	0.3954	0.4038
무게중심	0.3665	0.3738	0.3815	0.3897	0.3986
평탄도	0.3428	0.3512	0.3601	0.3696	0.3799

표 3. 부밴드 에너지 핑거프린트의 FRR (%)
Table 3. FRR of subband-energy fingerprint (%).

변환 \ FAR	10^{-20}	10^{-18}	10^{-16}	10^{-14}	10^{-12}
MP3 32 kbps	0	0	0	0	0
3dB EQ	0.12	0.10	0.05	0.02	0
LSC +1%	0.10	0.02	0.02	0	0
LSC -1%	0.17	0.07	0.05	0.02	0
TSM +1%	8.84	4.77	1.76	0.84	0.29
TSM -1%	9.93	5.71	2.58	0.92	0.34

표 4. 부밴드 무게중심 핑거프린트의 FRR (%)
Table 4. FRR of subband-centroid fingerprint (%).

변환 \ FAR	10^{-20}	10^{-18}	10^{-16}	10^{-14}	10^{-12}
MP3 32 kbps	2.36	1.86	1.28	0.80	0.36
3dB EQ	2	1.64	1.25	0.65	0.31
LSC +1%	3.49	2.39	1.49	0.80	0.27
LSC -1%	3.52	2.22	1.20	0.63	0.31
TSM +1%	41.6	36.8	29.3	20.3	11.2
TSM -1%	41.8	36.6	29.3	21.9	12.7

표 5. 부밴드 평탄도 핑거프린트의 FRR (%)
Table 5. FRR of subband-flatness fingerprint (%).

변환 \ FAR	10^{-20}	10^{-18}	10^{-16}	10^{-14}	10^{-12}
MP3 32 kbps	98.3	98.0	97.9	97.8	97.4
3dB EQ	99.9	99.9	99.9	99.8	99.7
LSC +1%	100	100	99.9	99.8	99.8
LSC -1%	100	99.9	99.9	99.8	99.8
TSM +1%	99.9	99.9	99.9	99.8	99.7
TSM -1%	100	99.9	99.9	99.9	99.7

프린트를 얻고, 원본 MP3 파일에서 얻은 특징과 비교를 수행하였다. 핑거프린트 비교에는 3.2장에서 기술한 바와 같이 10초의 오디오 세그먼트 (100개의 윈도우)를 사용하여 1600 비트의 해밍 거리를 이용하였다. 비교 결과는 각각의 특징 별로 표 2에 나온 문턱값을 사용하여 구하면 표 3, 4, 5와 같다. 예상했던데로 FAR과 FRR은 반비례 관계가 있음을 확인할 수 있고, 에너지와 무게중심에서 구한 핑거프린트는 대부분의 오디오 변형들에 대해서 어느 정도 이상의 강인성을 가짐을 알 수 있다. 반면에 평탄도 특징은 비록 [9]에서 사용한 방법으로 VQ (vector quantization)하여 핑거프린트를 만들 경우에 좋은 성능을 보였으나, 본 논문에서 제시한 방법으로 아주 작은 값의 FAR이 요구되는 상황에서는 이진핑거프린트를 만들어 사용하기에는 적합하지 않음을 알 수 있다.

기존의 압축도메인을 이용하지 않고 완전히 복호화한 후에 다시 오디오를 분석하여 핑거프린트를 추출하는 방법들과 [7,8] 비교하면, 특히 시간축 변환에 대해서 제안된 방법의 강인성이 기존의 비압축도메인 방법에 비해서 상대적으로 낮음을 알 수 있다. 이는 3.1절에서 기술한 바와 같이 MP3 프레임의 길이가 오디오 샘플링 주파수와 무관하게 576으로 고정되어 있어서 나타나는 것으로 분석된다.

V. 결 론

본 논문에서는 압축도메인 특징을 이용한 오디오 핑거프린팅 방법을 제안하고 여러 종류의 특징들에 대해서 성능을 비교 분석하였다. 대부분의 오디오가 압축되어 저장되고 있는 상황에서 압축도메인을 직접 이용하는 것은 계산량 또는 시간이 제약되는 휴대용 단말, AOD 등의 환경에서 유용할 수 있다. 제안된 방법은 오디오 압축에 널리 쓰이고 있는 MDCT 도메인을 이용하였으며, MDCT 도메인을 부밴드로 나누고 대표적인 모멘트 특징인 에너지, 무게중심, 평탄도로 부터 각각 핑거프린트를 얻었다. 각각의 얻어진 핑거프린트에 대해서 핑거프린트가 갖추어야할 네 가지 요소인 차별성, 강인성, 간결성, 계산용의성에 대해서 정성적, 정량적으로 분석하였다. 분석 결과 부밴드 에너지를 이용한 핑거프린트가 가장 우수한 성능을 보였으며, 부밴드 무게중심이 근접한 성능을 가짐을 알 수 있다. 그에 반해서 부밴드 평탄도를 이용한 핑거프린트의 성능은 다른 방법들에 비해서 크게 성능이 낮았다. 본 연구 결과를 통해서 오디오 압축도메인에서 직접 신호처리를 하여 의미있는 특징을 추출할 수 있음을 확인하였다. 따라서 본 연구 결과를 확장하여 장르 분류 등 다른 오디오 정보처리 문제에도 적용할 수 있을 것으로 기대된다.

감사의 글

본 연구는 지식경제부, 문화체육관광부 및 정보통신연구진흥원의 IT산업원천기술개발사업의 일환으로 수행하였음 [2009-S-017-01, 사용자 중심의 콘텐츠 보호·유통 기술 개발].

참고문헌

1. M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes,

M.Slaney, "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668-696, 2008.
2. 윤원중, 이강규, 박규식, "음악 정보검색 시스템을 위한 효율적인 특징 벡터 추출에 관한 연구," *한국음향학회지*, 23권, 7호, 532-539쪽, 2004.
3. 한학용, 허강인, 김수훈, "오디오 데이터의 특징 파라메터 구성에 따른 내용기반 분석," *한국음향학회지*, 21권 2호, 182-189쪽, 2002.
4. P. Cano, E. Battle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," *Journal of VLSI Signal Processing*, vol. 41, no. 3, pp. 271-284, 2005.
5. M. K. Mandal, F. Idris, and S. Panchanathan, "A critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing*, vol. 17, no. 7, pp. 513-529, 1999.
6. A. Spanias, T. Painter, and V. Atti, *Audio signal processing and coding*, Wiley, 2007.
7. J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *Proc. International Conf. on Music Information Retrieval*, pp. 144-148, Oct. 2002.
8. Jin S. Seo, Minho Jin, Sunil Lee, Dalwon Jang, Seungjae Lee, and C. D. Yoo, Audio fingerprinting based on normalized spectral subband centroids, in *Proc. IEEE ICASSP*, pp. 213-216, Mar. 2005.
9. J. Herre, E. Allamanche, and O. Hellumth, "Robust matching of audio signals using spectral flatness features," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 127-130, Oct. 2001.
10. K. Bradenburg, "MP3 and AAC explained", in *Proc. of the AES 17th Int. Conf. on high quality audio coding*, paper no. 99-110, Aug. 1999.
11. Y Wang, M Vilermo, and D Isherwood, "The impact of the relationship between MDCT and DFT on audio compression," in *Proc. IEEE Pacific-Rim Conference on Multimedia*, pp. 130-138, Dec. 2000.
12. Jin S. Seo, J.A. Haitsma, and T. Kalker, "Linear speed-change resilient audio fingerprinting," in *Proc. IEEE Benelux Workshop on MPCA*, pp. 45-48, Nov. 2002.

저자 약력

•서 진 수 (Jin Soo Seo)

1976년 4월 12일생
1998년 2월 : 한국과학기술원 전기 및 전자공학과 (공학사)
2000년 2월 : 한국과학기술원 전자전산학과 (공학석사)
2005년 2월 : 한국과학기술원 전자전산학과 (공학박사)
2005년 3월~2006년 2월 : 한국과학기술원 정보전자연구소 BK21 연구원
2006년 3월~2008년 2월 : 한국전자통신연구원 디지털콘텐츠 연구단 선임연구원
2008년 3월~ 현재 : 강릉원주대학교 전자공학과 조교수

•이 승 재 (Seungjae Lee)

1977년 10월 28일생
2003년 2월 : 연세대학교 전자공학과 (공학사)
2005년 2월 : 한국과학기술원 전자전산학과 (공학석사)
2005년 2월~ 현재 : 한국전자통신연구원 SW 콘텐츠 연구부문 선임연구원