

강인한 오디오 핑거프린팅 시스템을 위한 에너지와 통계적 필터링

Energy and Statistical Filtering for a Robust Audio Fingerprinting System

정병준, 김대진
(주)다이렉트미디어

Byeong-Jun Jeong(bjjeong@directmedia.net), Dae-Jin Kim(djkim@directmedia.net)

요약

디지털 음악과 스마트 폰이 대중화되면서 잡음에 강인한 실시간 음악 핑거프린트 시스템이 다양하게 개발되고 있다. 특히 핑거프린트 알고리즘 중 Multiple Hashing(MLH)은 잡음에 강인하고 정교한 구조로 되어 있다. 본 논문에서는 음악 데이터베이스로부터 질의 및 응답의 정확도를 개선하기 위해 에너지 집중 필터를 사용하고 연속성과 중복성을 제거하는 통계적 필터를 제안한다. 에너지 집중 필터는 하위 비트에 에너지가 집중되는 Discrete Cosine Transform(DCT)의 특징을 이용하고, 통계적 필터는 검색된 핑거프린트 정보들 사이의 상관관계 특징을 이용한다. 실험 결과로 잡음 환경에서 에너지와 통계적 필터링으로 구성된 제안 알고리즘은 우수성을 보인다. 이는 제안된 필터 엔진으로 Philips Robust Hash(PRH)보다 잡음에 강인하고 Multiple Hashing(MLH)보다 간결한 핑거프린트 시스템을 구성할 수 있다.

■ **중심어** : | 음악 검색 | 핑거프린트 | 에너지 집중 필터 | 연속성 제거 | 중복성 제거 |

Abstract

The popularity of digital music and smart phones led to develop noise-robust real-time audio fingerprinting system in various ways. In particular, The Multiple Hashing(MLH) of fingerprint algorithms is robust to noise and has an elaborate structure. In this paper, we propose a filter engine based on MLH to achieve better performance. In this approach, we compose an energy-intensive filter to improve the accuracy of Q/R from music database and a statistic filter to remove continuity and redundancy. The energy-intensive filter uses the Discrete Cosine Transform(DCT)'s feature gathering energy to low-order bits and the statistic filters use the correlation between searched fingerprint's information. Experimental results show that the superiority of proposed algorithm consists of the energy and statistical filtering in noise environment. It is found that the proposed filter engine achieves more robust to noise than Philips Robust Hash(PRH), and a more compact way than MLH.

■ **keyword** : | Music Retrieval | Fingerprint | Energy-intensive Filter | Remove Continuity | Remove Redundancy |

I. 서론

최근의 내용 기반 검색 서비스는 음악 특징을 추출하

는 방식인 음악 신호 분석에 의한 핑거프린트 기술을 이용한다. 핑거프린트(Fingerprint)를 이용하는 음악 인식 기술은 모바일 및 웹 등에서 많이 사용되고 있다. 모

바일이나 PC의 마이크를 이용하여 실시간으로 녹음한 음악에 대해 음악의 정보 및 콘텐츠의 다운로드 서비스를 제공하거나 유사 음악이나 동일 음악가의 음악을 서비스로서 제공한다. 음악 분류의 기준은 각 음악이 가지는 고유한 특징을 찾아 데이터베이스에 분류 기준을 만들고 새로운 음악이나 녹음된 음악의 데이터 특성을 기준으로 매칭되는 음악 정보를 제공하게 된다. 음악 인식 기술에서 사용되는 음악은 잡음 및 왜곡 등의 음악 훼손이 발생하는데 이런 조건에서 효과적이고 강인한 음악 신호의 특징을 찾아내어 음악을 검색하게 된다. 또한, 빠른 검색 결과를 보기 위해서는 실시간 검색이 가능해야 한다[1][2].

본 논문은 실시간으로 주어지는 음악 검색을 위해 강인한 음악 특징을 추출하고 빠른 대용량 데이터베이스 검색을 하는 핑거프린트 시스템을 제안한다. 음악 신호 분석에 의한 핑거프린트 시스템은 인간의 지문과 같이 음악이 가지는 고유한 값을 음악의 특징으로 사용하여 데이터 압축 및 색인을 만들고, 이를 이용하여 대용량 데이터베이스의 음악에서 효율적으로 음악 검색을 하고자 한다. 기존의 음악 검색 기술 중에서 가장 널리 알려진 방법은 Philips Robust Hash(PRH) 알고리즘이다 [3]. PRH 알고리즘은 실험적 수치에 의해 만들어진 핑거프린트의 비트 에러를 분석을 하는 방식이기 때문에 비교 대상간의 잡음에 따라 성능 차이가 발생한다. 이를 개선하고자 제안되었던 방법으로 주파수 변환에 따른 각 밴드별 에너지를 구하여 주파수 별 핑거프린트를 만드는 다중 해싱 방법 (Multiple Hashing Method: MLH)도 제시되었다[4]. 그러나 이 방법은 이산 코사인 변환(Discrete Cosine Transform, DCT)을 통한 주파수 변환에 의하여 에너지가 저주파 영역에 집중되기 때문에 비슷한 핑거프린트로 Lookup Table(LUT)이 구성된다. 따라서 사전 검색 시 사용되는 핑거프린트 후보가 증가하여 결과적으로 인식 속도에 불리한 측면이 발생된다[5]. 이에 본 논문은 잡음에 강인한 오디오 특징을 만들기 위하여 DCT를 적용하여 오디오 핑거프린트를 만들고[6][7], 빠른 대용량 데이터베이스 검색을 위한 매칭 필터를 다중으로 구성하여 검색 및 매칭 속도를 향상하고자 한다.

II. 관련 연구

1. 해시 기반의 핑거프린트 알고리즘

음악 검색 기술 중에서 가장 널리 알려진 방법인 PRH 알고리즘은 핑거프린트의 비트 오류율 분석을 하는 방식이다. 핑거프린트 생성을 위해 입력 음악을 5512Hz로 다운샘플링(Down-Sampling)한 후, 선택한 주파수 대역인 300~2000Hz의 주파수 대역 내에서 0.371 초 길이를 갖는 해닝 윈도우(Hanning Window) 및 31/32의 중첩 요소(Overlap Factor)에 의해 가중되는 해시 워드를 구성한다. 해시 워드는 33개로 나누어진 대역으로부터 얻을 수 있다. 해시 기반의 핑거프린트 추출을 개념도로 구성하면 [그림 1]과 같다[3].

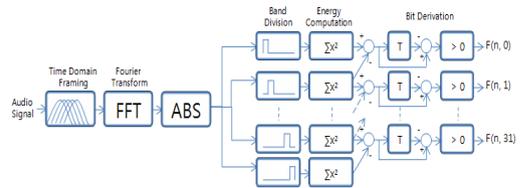


그림 1. 해시 기반의 핑거프린트 추출 개념도

2. 다중 해싱 기반의 핑거프린트 알고리즘

PRH 알고리즘이 잡음에 취약한 특성을 가지는 문제점을 개선하고자 다수의 해시 테이블을 생성하여 음악 검색 성능을 높이는 방법으로 다중 해싱 알고리즘이 제안되었다[4][8]. 주파수 계수 별로 LUT를 구성하기 때문에 효율적인 검색이 가능하도록 하였고, 주파수 계수 별로 해시 테이블을 생성하여 각 프레임마다 4개의 서브 핑거프린트가 만들어지도록 LUT를 구성하였고, 결과적으로 데이터베이스는 4개의 해시 테이블이 생성된다. 분할된 각 대역에서 잡음에 보다 강인한 특징을 구성하기 위해 DCT 방식의 특징 추출 방법을 사용한다 [4]. 다음 식(1~3)은 밴드별 에너지를 이용하여 서브 핑거프린트를 구성하는 방법을 나타낸다.

$$ED_k(n, m) = C_k(n, m) - C_k(n, m + 1) - (C_k(n - 1, m) - C_k(n - 1, m + 1)) \quad (1)$$

$$F_k(n, m) = \begin{cases} 1, & ED_k(n, m) > 0 \\ 0, & ED_k(n, m) \leq 0 \end{cases} \quad (2)$$

$$F_k(n) = [F_k(n, 0), \dots, F_k(n, 31)] \quad (3)$$

DCT 변환 후, 나뉜 서브밴드에서 1에서 32까지의 정수 길이 k를 갖는 연속된 서브밴드의 DCT 계수값을 각 밴드별 에너지인 $C_k(n, m)$ 로 표현한다. $C_k(n, m)$ 는 n번째 프레임의 m번째 서브밴드에서 구한다. 식(1)과 같이 프레임의 속성과 각 밴드의 속성을 포함하여 각 핑거프린트를 구성하기 위해 차분하여 구한 에너지 $ED_k(n, m)$ 을 구한다. $ED_k(n, m)$ 의 값을 결정적인 식(2)를 통해 n번째 프레임에서 구한 m번째 핑거프린트 비트인 $F_k(n, m)$ 를 구하여 최종적으로 식(3)과 같이 32개의 길이는 갖는 서브핑거프린트 $F_k(n)$ 를 구성한다.

[그림 2]와 같이 $4 * 32 * 256 = 32768$ 개 비트로 구성된 해시 워드 블록이 구성된다[2]. 구해진 해시 블록에 대해 해당되는 해시 테이블을 이용하여 PRH 방법과 동일하게 데이터베이스를 검색하고 비트 에러율(Bit Error Rate: BER)을 계산하여 비트 에러율이 임계값인 25% 이하인 경우 일치함을 판별해 낸다. 그러나 MLH 알고리즘은 검색의 정확도와 잡음의 강인성을 위하여 LUT가 4배 증가하여 대용량 음악에 대한 검색 시간이 증가하고, 데이터베이스 용량을 보다 많이 사용하기 때문에 대용량 음악 검색 서비스에 적합하지 않다.

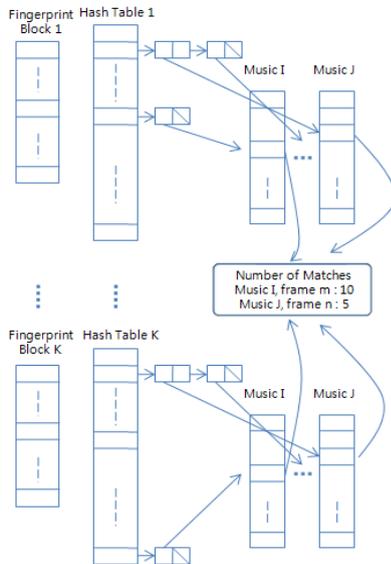


그림 2. 다중 해싱 기반의 핑거프린트 검색

III. 음악 검색 시스템 개선

1. 음악 검색 시스템 구성

본 논문에서는 MLH 방식과 동일한 핑거프린트 생성 과정을 수행한다. 생성된 해시 워드는 입력 오디오의 연속된 256개의 해시 워드들로 구성된 해시 블록을 생성한다. 이렇게 만들어진 해시 워드 및 해시 블록은 시간 축에 따라 높은 상관관계를 가진다. MLH와의 차이는 DCT 계수에 의해 다중으로 구성되는 LUT를 구성하지 않고 단일 LUT를 구성하여 MLH 방식의 단점인 검색 속도 저하를 보완하고자 하였다.

구성된 해시 블록으로 데이터베이스의 각 음악 해시 워드들을 검색하여 인식 결과를 도출한다. 검색해야 하는 대상의 수는 데이터베이스의 크기에 따라 선형적으로 증가한다. LUT는 32비트 해시 워드에 대한 위치인 엔트리(Entry)를 갖는다. 해시 워드가 다수 음악의 여러 위치들에 매칭되어 있을 수 있도록 노래에 대한 링크드 리스트(Linked list)로 저장된다. 매칭 방법은 단일 해시 워드가 비트 에러들을 갖지 않는다고 가정한다. 데이터베이스의 LUT에 존재하는지 판단하며 해시 블록 간 비교를 통해 검색이 이루어진다.



그림 3. 음악 검색 시스템 개선

제안하는 음악 검색 시스템 개선은 [그림 3]과 같은 순서로 수행된다. 3장에서 기술된 음악 검색 시스템에

의해 구성된 쿼리를 대상으로 중복성 제거 필터, 에너지 집중 필터와 연속성 제거 필터를 수행하여 시스템의 강인성과 검색 속도 개선을 한다.

2. 알고리즘 강인성 개선

2.1 에너지 집중 필터

에너지 집중은 음악 신호의 대역별 에너지를 구한 후 수행되는 DCT 연산의 특징에 의해 이루어진다. DCT 계수값으로 해시 블록이 만들어지면 DCT 특성에 의해 하위 비트에 에너지가 모이게 되는 특성을 이용한다. DCT 적용 예로 [그림 4]와 같이 입력된 음악 신호에 DCT를 수행했을 때 하위 비트에서 저주파 부분을 나타내고 상위 비트에서 고주파 부분을 나타내게 된다. DCT를 거친 시간 개념의 쿼리를 통해 고주파는 사용하지 않고 잡음에 강인한 저주파 부분의 비트만을 사용하면 검색 연산을 할 때 데이터 량이 감소하므로 제안하는 알고리즘의 이점이 생긴다. 본 논문에서는 DCT 에너지 집중도를 기준으로 부분 핑거프린트 워드 비트를 선택적으로 테스트하여 가장 신뢰도가 높은 일부 비트를 추출하는 방법을 사용하였다.

검출된 해시 워드 중에서 DC 성분을 제외하고 저주파 영역에서 해당하는 25번째의 주파수 계수값을 통하여 비트 에러율을 구하게 된다. 이때 임계값은 실험적으로 0.15보다 낮으면 동일 음악으로 식별된다.

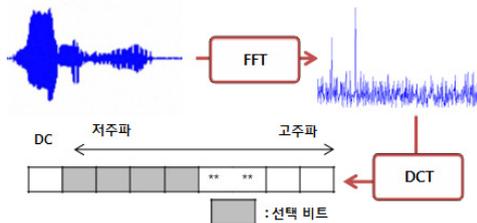


그림 4. 에너지 집중 필터의 비트 선택 검출 흐름도

3. 핑거프린트 검색 속도 개선

중첩된 프레임에 DCT를 하게 되면 하위 에너지 비트에 에너지가 집중되는 현상이 발생한다. 이와 같은 특징은 만들어진 핑거프린트 해시 워드가 데이터베이스 검색 시에 동일 음악의 인접 위치에 다수 분포하게

되는 특성을 나타내게 한다. 따라서 본 논문에서는 음원에서 근접 거리에 매칭되는 다수의 핑거프린트를 제거하는 중복성 제거 필터와 시간적으로 연속 분포하는 핑거프린트를 제거하는 연속성 제거 필터를 구성하여 검색 속도 개선을 한다.

3.1 중복성 제거 필터

후보로 검색되는 핑거프린트는 LUT의 구성되는 특성에 의해 다수의 개수를 가질 수 있다. 하나의 핑거프린트에 해당하는 검색 연산은 해시 블록 생성과 비트 에러율 계산과 같이 많은 연산량을 필요로 한다. 생성된 핑거프린트 리스트 중에 동일 음악에서 생성된 핑거프린트가 동일 해시 블록에 존재할 때, 불필요한 연산을 하게 된다. 따라서 이런 중복 검색의 불필요성을 제거하기 위해 중복성 제거 필터를 구성하였다.

MLH 알고리즘의 경우 해시 블록 중에 비트 에러율이 25%이내면 검색이 되도록 정의하였다[4]. 즉, 해시 블록은 256개의 해시 워드로 구성되면, 1개의 해시 워드는 32비트로 구성되고 총 8192(256 * 32)의 비트 중 2048개 이상의 비트가 같으면 음악 검색이 되었다. 이것은 해시 블록의 256개에서 약 60개의 해시 워드가 모두 비트 에러가 발생해도 정상 범위 안에서 음악 검색이 되는 것이다.

만약, LUT에서 검색된 해시 워드가 [그림 5]처럼 해시 워드 0x10101010과 같고, LUT에서 검색된 링크드 리스트 값 중에 음악 번호가 1, 위치 번호가 9998인 것과 음악 번호가 1, 위치 번호가 10000인 것은 [그림 5]의 (A), (B)와 같이 대부분 동일한 부분(회색 영역)에 대하여 비트 에러율을 계산한다. 임의의 위치 번호 X에서 만들어진 해시 블록과 쿼리 해시 블록 사이의 비트 에러율이 0이라는 최상의 상황을 가정하면, 동일한 음악 번호를 갖는 위치 번호 X - 60부터 X + 60까지에서 각 위치 번호를 기준으로 만들어진 해시 블록과 쿼리 해시 블록 사이의 비트 에러율들은 MLH에서 설정한 비트 에러율 임계치(허용 비트 에러율 25%, 해시워드 단위 60개 차이 이내)보다 낮은 비트 에러율이 된다. 즉, 위치 번호가 MLH 알고리즘에서 허용한 비트 에러율 25%에 해당하는 60개의 위치 번호 차이 이내라면, 중복

하여 해시 블록의 비트 에러율을 계산한 것과 같다. 따라서 본 논문에서는 위치 번호가 일정 거리 차이 이내의 경우에 중복성 임계치(Continuity Threshold)를 설정하여 중복성 제거 필터를 구성한다.

LUT에서 해시 블록의 n번째 쿼리 Q_n 과 Q_n 에 의해 찾아진 K개의 음악번호 리스트 ID, 음악위치 번호 리스트 POS는 $L(Q_n, ID_k, POS_k), k = 1, 2, \dots, K-1$ 이 된다. 이와 같이 정의하면 중복성 제거 필터는 다음의 의사 코드(Pseudo Code)로 수행된다.

```

for k = 0 to K-1
  for k' = k + 1 to K-1
    if IDk == IDk'
      Calculate 1-norm Distance
        between POSk and POSk'
      if 1-norm Distance < Continuity Threshold
        Delete POSk'
      end
    end
  end
end
end
end
    
```

본 논문에서는 위치 번호의 중복성 임계치를 실험적으로 얻어진 38을 사용한다. 테스트로 3000곡에 대하여 데이터베이스를 만들었을 때 중복성 제거 필터를 사용하는 음악은 78곡이 발생하였고 이때 정상 검색이 되었다.

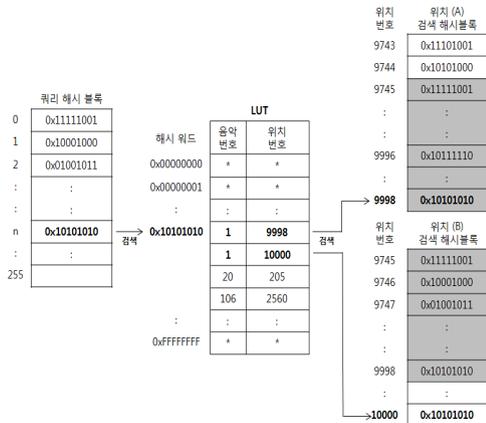


그림 5. 중복성 제거 필터

3.2 연속성 제거 필터

비트 에러율을 검사한 결과 임계값보다 적은 값을 가지는 해시 블록을 질의에 해당하는 음악으로 검색하게 된다. 그러나 질의로 들어오는 음악은 주변 상황에 따라 다양한 잡음 환경에 노출되고 있어서 원본과 차이가 발생하게 된다. 약간의 잡음이 지속적으로 원본을 손상시키게 되면 대부분 임계치보다 높게 되어 음악 검색을 못하는 결과가 생긴다. 테스트 중에 질의 음악이 검색하고자 하는 음악이지만 임계치를 조심씩 넘어서면서 연속적으로 똑같은 음악 번호를 가지는 경우가 발생한다. 본 논문에서는 검색 속도를 높이기 위하여 이와 같은 경우에 검색하고자 하는 콘텐츠로 판정한다.

LUT에서 해시 블록의 n번째 쿼리 Q_n 과 Q_n 에 의해 찾아진 K개의 음악번호 리스트 ID, 음악위치 번호 리스트 POS는 $L(Q_n, ID_k, POS_k), k = 1, 2, \dots, K-1$ 이 된다. 이를 의사 코드로 표현하면 다음과 같다.

```

if Qn's BER < BER Threshold
  for k = 0 to K-1
    for k' = 0 to K-1
      if IDk == IDk'
        Find Maximum of POSk'(POSmax)
        Find Minimum of POSk'(POSmin)
        Count POSk'(POScnt)
      end
      if POScnt > Recundancy Threshold
        and POScnt == POSmax - POSmin
          Return Find Music ID = IDk
        end
      end
    end
  end
end
    
```

핑거프린트 블록의 비트 에러율 임계치(BER Threshold)설정은 MLH 방법에서 사용되는 25%보다 큰 35%를 실험적으로 구하여 사용한다. 동일 음악에서 연속적으로 연속성 임계치(Recundancy Threshold)인 15개 이상의 핑거프린트가 검색되는 경우 검색한다.

연속성 제거 필터를 테스트하기 위해 3000곡에 대하여 데이터베이스를 만들고 핑거프린트가 연속적으로 구성된 경우에 해당 테스트를 통해 해당 음악을 검색하였다. 질의 음악을 테스트하여 연속성 제거 필터에 의해 검색되는 경우가 98곡 존재 하였으며, 해당 음악을 오인식하는 경우는 존재하지 않았다. 대용량의 음악에 대한 경우, 본 실험을 통해 얻은 결과와 달리 오인식될 경우가 증가할 수 있고 정확도 측면에서도 낮은 성능을 보일 수 있지만, 음악 검색 서비스의 목적에 맞게 검색 속도의 성능을 높이는 부분에서 좋은 결과를 얻을 수 있다.

IV. 실험

1. 환경

제안하는 알고리즘의 음악 검색 성능을 판단하기 위해 테스트 환경을 설정하였다[9][10]. 데이터베이스는 테스트 대상 곡을 포함하여 20개 장르의 3000곡으로 구성하였고, 테스트 대상 음악은 다음과 같이 음악 240곡에 대하여 녹음 및 필터링된 21가지 환경(총 5040곡)으로 음악 검색 테스트를 하였다.

- ① White Noise : 0~20dB SNR (Step : 5dB) 추가
- ② Split Time : mp3에서 40~140초 구간 선택
- ③ Time Scale Modification : 피치의 변화 없이 템포를 10% 증가
- ④ Recording : 각 장소 별 녹음(Silence Office, Noise Office, Main Street, Back Street, Café , Department)
- ⑤ Echo : 원 신호 대비 90% 에코 및 0.1초 응답 구성
- ⑥ Equalization : [표 1]과 같이 10개 밴드 형태로 파라미터 선택
- ⑦ Transcoding : mp3에서 AAC로 변환 구성
- ⑧ Down Sampling : Mp3의 인코딩/디코딩을 통해 32Kbps, 128Kbps 구성
- ⑨ Band Pass : 주파수 필터링을 통해 100 ~ 6000Hz 대역만 선택

- ⑩ Linear Speed Change : 템포와 피치 변화를 기준으로 -4%, +4% 변환

표 1. Equalization 밴드 구성

Freq.(Hz)	31	62	125	250	500	1k	2k	4k	8k	16k
Gain(dB)	-3	+3	-3	+3	-3	+3	-3	+3	-3	+3

2. 실험 결과

잡음 환경에서 강인성을 가지기 위해 에너지 집중도가 높은 구간인 하위 25비트 구간을 선택하여 이 구간의 해당하는 해시 워드를 데이터베이스에 추가 저장하고, 이 값을 음악 검색에 사용하였다. 또한 연속성 제거 필터 기반의 기능을 추가하여 매칭되는 해시 워드가 연속하여 15개 이상 발생하는 경우 질의 음악에 대한 검색 결과를 나타내었다. 또한, 입력된 단일 해시 워드부터 데이터베이스의 검색된 해시 워드들이 동일 음악 이면서 해시 워드의 거리가 38 이내의 위치에 해시 워드가 존재하는 경우 중복성 제거 필터를 사용하였다.

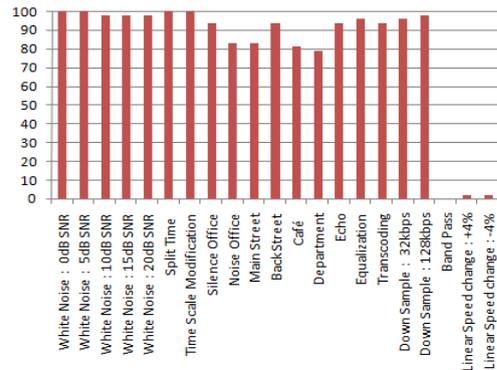


그림 6. 환경 별 제안 알고리즘의 검색 성능

[그림 6]과 같이 핑거프린트에 DCT 특성을 사용함으로써 잡음 환경에서 강인한 특성을 보였으며, 길거리와 카페 및 공공 장소에서 녹음된 음악 테스트 등에서 발생하는 인위적, 자연 발생적 화이트노이즈 잡음에 강인한 특성을 보임을 알 수 있다. 또한 에코 추가와 평활화, 포맷 변환으로 설정한 테스트 환경에서도 우수한 성능을 보임으로서 본 논문에서 추구하고자 하는 서비스 형태의 음악 검색에 적합한 성능을 보인다. 하지만, 시간

순서에 기반한 해시 워드 생성으로 인하여 시간적 요소에 의한 실험 테스트인 시변환 필터 적용과 인각 청각 시스템 기반의 특정 대역을 사용함으로써 특정 대역만을 사용하는 필터에서 다소 낮은 성능을 보였다. PRH, MLH 또한, 대부분의 시간 순서에 기반으로 한 핑거프린트에서는 거의 검색이 되지 않는다. 그러나 이런 특성은 실험으로서 파악된 부분이며, 음악 검색 서비스에서 발생할 확률이 낮은 특성으로 음악 서비스라는 부분에서는 큰 영향을 줄 수 있는 특성은 아니다.

V. 성능 비교

다음 [표 2]는 제안하는 알고리즘과 PRH, MLH 알고리즘을 여러 환경에서 녹음된 음악을 대상으로 하는 검색 테스트 비교 결과이다.

표 2. 성능 비교

환경	제안 알고리즘	PRH	MLH
White Noise : 0dB SNR	100	100	100
White Noise : 5dB SNR	100	100	100
White Noise : 10dB SNR	97.91	97.91	97.91
White Noise : 15dB SNR	97.91	93.75	95.83
White Noise : 20dB SNR	97.91	87.5	91.66
Split Time	100	100	100
Time Scale Modification	100	100	100
Silence Office	93.75	85.42	93.75
Noise Office	83.33	62.5	77.08
Main Street	83.33	62.66	77.08
Back Street	93.75	83.33	87.5
Caf	81.25	64.58	81.25
Department	79.16	62.5	77.08
Echo	93.75	89.58	93.75
Equalization	95.83	91.66	91.66
Transcoding	93.75	91.66	93.75
Down Sample : 32kbps	95.83	89.58	91.66
Down Sample : 128kbps	97.91	91.66	93.75
Band Pass	0	0	0
Linear Speed change : +4%	2.08	0	0
Linear Speed change : -4%	2.08	0	0

표 3. 알고리즘별 FAR 결과

알고리즘	제안 알고리즘	PRH	MLH
FAR(%)	7	5	19

제안 알고리즘은 PRH, MLH와 비교하여 필터가 추가된 환경 테스트 결과에서 전체적으로 우수한 성능을 보이며 녹음 곡에서도 일반 환경인 실내, 공공장소 등의 결과를 기준으로 더 우수한 성능을 보인다. 그러나 제안 알고리즘, PRH와 MLH는 해시워드 생성 방법이 시간 순에 의한 것이기 때문에 시간 개념의 필터가 추가하는 실험 환경인 대역 통과, 선형 시간 변형 테스트 곡에서 모두 낮은 성능을 보인다. 검색 정확도에서 제안 알고리즘은 DCT를 사용함으로써 잡음 환경에 강인한 특성을 보이고 에너지 집중 필터를 통해 한 번 더 해시 블록의 비트 에러율을 검사함으로써 검색의 정확도를 높였기 때문에 PRH, MLH에 비하여 상대적으로 우수한 성능을 보인다.

검색 속도에 있어 제안 알고리즘은 PRH 알고리즘과 동일한 크기의 LUT 테이블을 구성하기 때문에 유사한 성능을 보인다. 하지만, MLH 알고리즘 방식은 일반적으로 LUT 테이블을 4개 사용하기 때문에 데이터베이스 구성에 있어 제안 알고리즘보다 비트 에러율 측정 시 약 4배 이상의 시간이 소요되었다. 또한 MLH는 DCT 계수값을 핑거프린트로 사용하기 때문에 저주파로 에너지가 집중되는 특징으로 인해 LUT로부터 질의 검색 시 PRH보다 많은 수의 모집단을 구성한다. 따라서 비트 에러율을 계산해야 할 대상이 많아 대용량 서비스에 적합하지 않다.

3000곡 테스트 시 [표 3]과 같이 MLH는 상대적으로 데이터베이스에 존재하는 다른 노래로 인식하는 오인식(False Acceptance Rate: FAR)이 높게 발생하였으나, 제안하는 알고리즘은 PRH와 유사하게 FAR이 상대적으로 적게 발생하였다. 제안하는 알고리즘에서도 DCT를 사용하기 때문에 잡음 환경에 강인하나 다수의 모집단을 구성하게 되어 연산 속도가 증가할 수 있다.

표 4. 알고리즘 적용에 따른 속도 비교

알고리즘	제안 알고리즘 적용 전	제안 알고리즘 적용 후
곡 당 인식 평균 소요(초)	5.242	1.578

본 논문에서는 속도의 개선을 위해서 중복성 제거 필터와 연속성 제거 필터를 사용하여 [표 4]와 같이 기존 알고리즘 대비 약 3.3배의 음악 검색 속도 개선을 하였다. 또한 실험 결과들과 같이 다양한 테스트를 통해 음악 검색 서비스에 제안 알고리즘이 PRH나 MLH보다 적합함을 확인하였다.

VI. 결론

본 논문에서 제안하는 알고리즘은 일반적인 환경에서 취득되는 음악을 대상으로 해시 워드의 특징 추출 측면에서 강한 특성을 보이고, 음악 검색 서비스에 적용하기 위하여 검색 결과 응답이 실시간 검색 기능을 확보하는 방법을 제시하고자 한다. 음악의 특성을 추출하는 방식에서 해시 워드를 사용하는 PRH 방식의 단점인 잡음의 취약한 점과 이의 대안으로 제시된 MLH 방식에서 LUT 테이블의 증가로 인한 단점인 시스템의 간결성 부족을 개선하고자 하였다. 음악 특성 추출은 MLH 방식과 동일하게 FFT와 DCT를 사용하여 잡음에 강한 특성을 가진 핑거프린트들을 구성하였다. 음악 검색 알고리즘은 에너지 집중 필터를 추가 구성하여 검색 정확도를 높였고, 시간적으로 연속하여 존재하는 해시 워드에 대하여 연속성 제거 필터와 음악 내에 인접 위치에 분포하는 해시 워드의 집중을 제거하는 중복성 제거 필터를 구성하여 음악 검색 서비스에 적용함으로써 검색 속도를 높였다.

참고 문헌

- [1] J. Haitsma and T. Kalker, "Speed-change resistant audio fingerprinting using auto-correlation," *Acoustics, Speech and Signal Processing*, 2003.
- [2] 서용석, 김원겸, 이선화, 서영호, 황치정, "이미지 콘텐츠 출력물의 저작권보호를 위한 디지털 핑거프린팅 기술에 관한 연구", *한국콘텐츠학회 추계 종합학술대회 논문집*, 제4권, 제2호, 2006(11).
- [3] J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System," *Proc. 3rd Int. Conf. Music Information Retrieval*, pp.107-115, 2002(10).
- [4] Yu Liu, H. S. Yun, J. S. Sung, and N. S. Kim, "A Novel Audio Fingerprinting Scheme based on Subband Envelop Hashing," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp.813-816, 2009(10).
- [5] P. Doets and R. Lagendijk, "Distortion estimation in compressed music using only audio fingerprints," *IEEE Trans. Audio, Speech, and Language Processing*, Vol.16, No.2, pp.302-317, 2008(2).
- [6] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Computers*, pp.90-93, 1974(1).
- [7] K. Rao and P. Yip, "*Discrete Cosine Transform: Algorithms, Advantages, Applications*," Academic Press, 1990.
- [8] Jianping Chen and Tiejun Huang, "A Robust Feature Extraction Algorithm for Audio Fingerprinting," *Computer Science, advances in multimedia information processing*, Vol.5353/2008, pp.887-890, 2008.
- [9] Avery Li-Chun Wang, "An Industrial-Strength Audio Search Algorithm," *4th Symposium Conference on Music Information Retrieval*, pp.7-13, 2003(10).
- [10] Cerling Wold, Thom Blum, Douglas Keislar, and James Wheaton, "Content-Based Classification, Search, and Retrieval of Audio," *IEEE*

[1] J. Haitsma and T. Kalker, "Speed-change resistant audio fingerprinting using auto-

Multimedia, Vol.3, No.3, pp.27-36, Fall, 1996.

저 자 소 개

정 병 준(Byeong-Jun Jeong)

정회원



- 2007년 2월 : 인천대학교 정보통신공학과 졸업(공학사)
- 2009년 2월 : 인천대학교 정보통신공학과 졸업(공학석사)
- 2009년 ~ 2011년 : 로봇에버 기술연구소 주임연구원

▪ 2011년 ~ 현재 : 다이렉트미디어 기술연구소 연구원
 <관심분야> : 디지털 콘텐츠, 내용기반 검색, 정보 보안, 저작권 보호, 패턴 인식

김 대 진(Dae-Jin Kim)

정회원



- 1998년 2월 : 대전대학교 전자공학과 졸업(공학사)
- 2000년 2월 : 동국대학교 전자공학과 졸업(공학석사)
- 2010년 2월 : 대전대학교 전자공학과 졸업(공학박사)

▪ 2000년 ~ 2003년 : 한빛소프트 주임연구원
 ▪ 2003년 ~ 2007년 : 모토로라 코리아 전임연구원
 ▪ 2007년 ~ 2008년 : 아이비인터넷 부장
 ▪ 2008년 ~ 2010년 : 피어컴(미디어웹) 선임연구원
 ▪ 2010년 ~ 현재 : 다이렉트미디어 기술연구소 연구소장
 <관심분야> : 저작권 보호, 멀티미디어 시스템, 디지털 콘텐츠, 멀티미디어 검색, IPTV, 오디오 핑거프린트, 비디오 카툰닝 등