

논문 2008-45SP-6-18

잡음에 강인한 내용기반 음악 검색 시스템에 대한 연구

(A Study of Noise Robust Content-Based Music Retrieval System)

윤원중*, 박규식*

(Won-Jung Yoon and Kyu-Sik Park)

요약

본 논문에서는 모바일 환경에서 적용 가능한 잡음에 강인한 내용기반 음악 검색 시스템을 구축하였다. 제안된 시스템은 기존의 음성인식 분야에서 잡음에 강인한 특성을 가진 것으로 알려진 ZCPA 특징을 내용기반 음악 검색 시스템에 적용시켜 그 성능을 검증하였다. 또한 본 논문에서는 대용량 음악 DB 검색에서 기존의 전수(Exhaustive) 검색에 비해 검색 속도를 99% 가까이 개선할 수 있는 새로운 인덱싱 방법과 고속 검색 알고리즘을 제안하였다. 신호대 잡음비가 15dB - 0dB인 잡음 환경에서의 모의실험 결과, 제안 시스템은 기존의 MFCC와 필터뱅크 에너지 특징에 비해 약 5% - 30% 정도의 우수한 성능을 나타냄을 확인하였다.

Abstract

In this paper, we constructed the noise robust content-based music retrieval system in mobile environment. The performance of the proposed system was verified with ZCPA feature which is known to have noise robust characteristic in speech recognition application. In addition, new indexing and fast retrieval method are proposed to improve retrieval speed about 99% compare to exhaustive retrieval for large music DB. From the computer simulation results in noise environment of 15dB - 0dB SNR, we confirm the superior performance of the proposed system about 5% - 30% compared to MFCC and FBE(filter bank energy) feature.

Keywords : content-based music retrieval, ZCPA, indexing method, fast retrieval method.

I. 서론

많은 멀티미디어 데이터들 가운데 음악은 인간의 삶과 밀접한 관계를 유지하면서 늘 인간과 함께하였다. 음악은 기쁨 때나 슬픔 때, 신앙심을 표현하기 위해서도 사용되었고, 스포츠 선수들에게는 동기 부여를 위해 사용되기도 하며, 최근에는 신경장애의 치료 또는 학습능력의 향상을 위해서도 사용되고 있다. 이렇듯 음악이 인간과 함께하며 경제의 균형 및 발전을 위하여 중요한

역할을 한다는 것은 기정사실로써 2005년도 음악 시장 규모는 3,352억원이었으며, 이중 디지털 음원 산업은 2,014억원으로 전체의 60%에 다다르고 있다^[1]. 그러나 이러한 디지털 음악 산업의 성장에도 불구하고 음악 포털사이트 등의 음악 DB에서 사용자가 원하는 음악을 효율적으로 찾을 수 있는 방법은 별로 없었다. 현재까지 음악 DB에서의 검색은 ID3 태그 또는 가수, 제목, 장르와 같은 카테고리에 근거하는 텍스트 검색 방식이었다. 반면, 내용기반 음악 검색 시스템은 음악의 내용을 수학적으로 분석하여 음악 DB에서 검색을 하는 방법으로, 드라마나 영화 또는 CF등을 시청하다가 스피커에서 흘러나오는 삽입곡, OST 등을 휴대폰의 마이크로폰을 이용하여 검색할 수 있다. 최근 국내 음악 시장에는 블로그와 미니홈피에 배경음악(BGM)을 삽입하는 서비스가 선풍적인 인기를 끌고 있는데, 웹 서핑을 하

* 정희원, 단국대학교 컴퓨터과학 및 통계학과
(Dept. of Computer Science and Statistics,
Dankook University)

※ 본 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.
(KRF-2007-511-D00197)

접수일자: 2008년5월19일, 수정완료일: 2008년10월15일

던 중 우연히 접속한 홈페이지에서 흘러나오는 좋은 음악의 제목을 모른다면, 사운드카드를 이용하여 그 음악의 일부를 바로 녹음한 후 음악 포털사이트를 통하여 검색, 구매할 수 있다. 또한, 작곡가들은 이 기술을 이용하여 누군가가 자신들의 음악을 표절하고 있는지 또는 그들의 새로운 음악이 표절 시비에 휩싸일지를 쉽게 알아보고 저작권 문제를 해결하거나 피해갈 수도 있으며, 음원 저작권자들은 TV나 라디오를 모니터링하여 자신들의 음악이 어느 정도 방송되어서 얼마만큼의 저작권료를 방송사에 요구할 수 있는가에 대한 근거자료를 구축할 수도 있다.

이전의 내용기반 음악 검색 시스템에 관련된 연구들^[2-9]은 PC나 웹 같이 잡음이 없는 환경에서의 연구가 주를 이루었다. 그러나 핸드폰이 생활필수품이 되어버린 현재, 모바일 환경에서 실제 적용 가능한 검색 시스템에 대한 요구가 점차 늘고 있는 실정이다.

Phillips^[10]는 잡음과 각종 신호적 변형(압축 등)에 강인한 특징을 추출하였다. 음악 신호의 분석 주파수 범위를 인간의 청각 시스템(Human Auditory System)과 밀접한 300-3000Hz의 주파수 범위로 제한하여, 약 3초의 음악 신호에 대하여 256개의 해시 값들로 구성된 1개의 해시 블록을 생성시킨 후, 이것을 음악 검색을 위한 특징으로 사용한다.

Shazam^[11] 역시 분석 주파수 범위를 300-3500Hz로 제한하여 각 프레임에서 지역 최대값(local maxima)를 찾아 핑거프린트로 사용한다. 각 핑거프린트의 타임포인트를 랜드마크(landmark)라 하여 핑거프린트와 함께 저장하는 방법을 제안하였다. 검색시 1차적으로 질의 음악과 동일한 핑거프린트들을 모두 선별한 후 각 데이터들 간의 선형적인 위치관계를 확인하여 최종 선택하게 된다. 그러나 위의 두 시스템 역시 잡음이 심한 상황(0dB~15dB)에서의 성능은 그리 만족스럽지 못한 실정이다.

이에 본 논문에서는 기존의 음성인식 분야에서 잡음에 강인한 특성을 가진 것으로 알려져 있는 ZCPA (Zero-Crossing with Peak Amplitudes)^[12] 특징을 내용기반 음악 검색 시스템에 적용시켜 그 성능을 검증하였다. 아울러 실제 대용량 음악 DB에서 적용 가능한 인덱싱 방법과 고속 검색 알고리즘에 대한 연구를 통하여 모바일 환경에서 적용 가능한 음악 검색 시스템을 구축하였다.

본 논문의 구성은 다음과 같다. 먼저 II장에서는 제

안된 시스템의 구조에 대하여 알아보고, III장에서는 ZCPA특징에 대하여 간략하게 설명하였다. IV장에서는 컴퓨터 모의실험을 통하여 제안된 시스템의 성능을 확인하고, 마지막으로 결론으로 끝을 맺는다.

II. 제안 시스템의 구조

다음 그림 1은 본 논문에서 제안된 내용기반 음악 검색 시스템의 구조를 나타낸다.

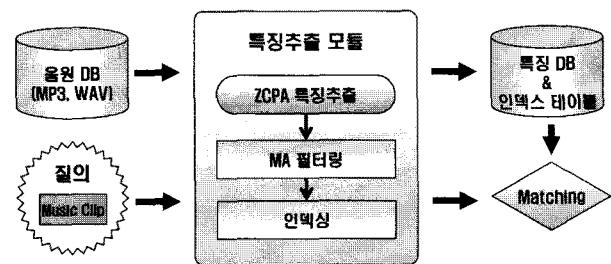


그림 1. 제안 시스템의 구조도
Fig. 1. Proposed system architecture.

우선 음원 DB에 있는 음악들에서 잡음에 강인한 ZCPA 특징을 추출하여 검색을 위한 특징 DB에 저장하고, 인덱스 검색을 위한 테이블을 생성하여 함께 저장한다. 인덱스 테이블은 ZCPA 특징을 추출한 후 MA(Moving Average) 필터링 과정을 거쳐 생성하게 된다. 이렇게 검색 서버에 특징 DB와 인덱스 테이블을 생성한 후, 미지의(Unknown) 음악 클립이나 녹음된 음악 신호가 질의로 서버에 받아들여지면, DB 구축시와 동일한 방법으로 ZCPA 특징과 인덱스를 추출하여 DB 내의 ZCPA 특징들과의 유사도를 측정한다. 검색 결과를 사용자에게 알려주게 된다. 특징 추출과 검색에 대한 자세한 설명은 다음 장에 기술하였다.

III. 특징 추출

최근 몇 년간 내용기반 음악 검색 분야에서는 음악을 효과적으로 분석할 수 있는 특징들을 추출하려는 노력이 계속되어왔고, 일반적으로 음성 신호처리 연구에서 많이 사용되어 왔던 필터뱅크 에너지(Filter-Bank Energy)^[10], Fourier 계수^[13], MFCC^[14], spectral flatness^[15], sharpness^[15], LPC 계수^[15] 등이 음악 검색을 위해 주로 사용되고 있다. 그러나 이들 특징들은 비교적 잡음이 없는 환경에서는 어느 정도 만족할 만한 성능을 보여주지만 배경 잡음이 존재하는 경우 특징 추출 과정

에서 왜곡된 특징벡터가 생성되고 이로 인해 성능이 저하되게 된다.

한편 인간은 잡음이 존재하는 환경에서도 매우 뛰어난 지각 능력을 보이고 있으며, 이러한 인간의 청각 시스템을 모델링함으로써 잡음에 강인한 특징 추출 방법을 개발하고자 하는 연구가 꾸준히 계속되고 있다. 이러한 청각 시스템을 모델링하고자 하는 연구의 하나로 ZCPA 모델^[12]이 제안되었다. ZCPA는 인간의 청각 시스템 중 기저막(Basilar Membrane)과 청신경 섬유(Auditory Nerve Fiber)를 모델링한 것으로 특히 잡음 환경에서 우수한 성능을 보인다. 이 장에서는 ZCPA 특징을 추출하는 과정을 설명하도록 한다.

1. ZCPA 특징 추출^[12]

본 논문에서 사용하는 ZCPA 모델은 인간의 청각 특성을 모델링한 EIH(Ensemble Interval Histogram) 모델^[16]을 기반으로 하고 있다. 그림 2는 ZCPA 모델의 블록도이다.

ZCPA 모델에서는 각 대역 통과 필터를 거친 신호의 상향 영교차점(upward zero-crossing point)을 파악하여 인접한 두 영교차점간의 시간 간격을 측정하고, 그 역수를 주파수 히스토그램에 누적시킨다. 그리고 인접한 영교차점간의 신호의 피크 값을 검출하여 비선형 함수를 통과한 값을 주파수 히스토그램에 가중치로 사용한다. 최종 ZCPA의 출력은 모든 채널의 히스토그램을 더함으로써 얻어지며, n번째 프레임의 ZCPA의 출력은 다음 수식 (1)과 같이 표현할 수 있다.

$$y(n, i) = \sum_{channel k=1}^{K-1} \sum_{j_k} \delta_{ij_k} g(A_k), \quad 1 \leq i \leq M \quad (1)$$

여기서 K는 각 채널에서 상향 영교차점의 개수를, M

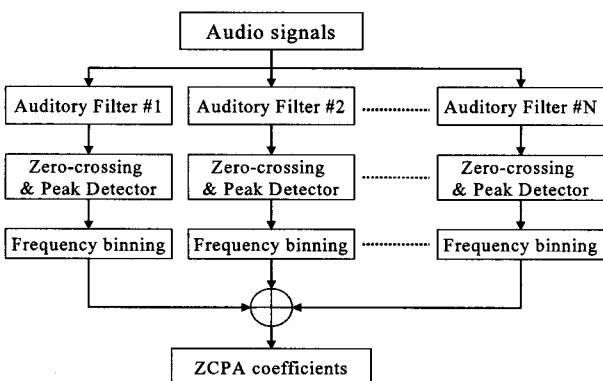


그림 2. ZCPA 모델의 블록도
Fig. 2. Block diagram of ZCPA model.

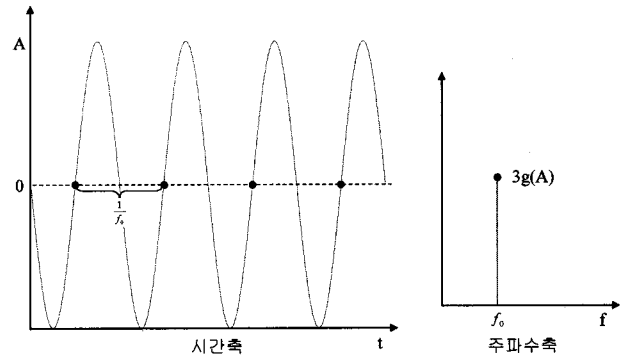


그림 3. ZCPA 동작의 한 예
Fig. 3. Example of ZCPA procedure.

은 주파수 히스토그램에서의 주파수 분할 구간의 개수를, j_k 는 k번째와 (k+1)번째 영교차점으로부터 계산된 주파수 구간의 인덱스(index)이다. 한편, A_k 는 k번째와 (k+1)번째 영교차점 사이 신호의 피크 값이고, δ_{ij} 는 Kronecker 델타함수, $g(\cdot)$ 는 단조증가(monotonic) 함수로 log 함수가 사용된다. ZCPA 동작의 이해를 돕기 위해 채널에 정현파 신호가 입력되었을 때의 한 예를 그림 3에 나타내었다.

그림 3의 시간축에서의 입력신호는 4개의 상향 영교차점을 갖고, 각 영교차점간의 시간간격은 $1/f_0$ 이기 때문에 이를 주파수 히스토그램으로 표현하면 f_0 지점에서 $3g(A)$ 의 크기로 표현된다. 또한, ZCPA 출력의 계산을 위해서는 각 채널마다 유한한 길이의 신호가 고려되어야 한다. 각 채널의 중심 주파수를 F_c 라고 할 때, 시간 t에서의 ZCPA 출력을 얻기 위해서는 구간 $[t-10/F_c, t]$ 만큼의 신호가 고려되는데 그 결과, 저주파 채널에서는 긴 신호가, 고주파 채널에서는 상대적으로 짧은 신호가 사용되므로 저주파 영역에서는 주파수 분해능(Frequency resolution)이 좋아지는 반면 시간 분해능이 나빠지게 되고, 고주파 영역에서는 이와 반대의 특성이 나타난다. 이는 실제 인간의 청각 특성과 일치하는 특성이다.

IV. 실험 및 결과 고찰

1. 실험 데이터

제안된 알고리즘의 검증을 위해 사용된 음악 데이터 베이스는 인터넷 전문 음악 사이트의 MP3 파일, 음악 CD 그리고 인터넷 라디오 방송으로부터 수집되었다. Classic, Hiphop, Jazz 그리고 Rock의 4개 음악 장르에 대하여 각 장르별로 500곡을 선정하여 총 2,000곡을 사용하였다. 또한 본 논문에서의 실험은 모바일 환경에서

의 검색 시스템을 고려하여 변화한 거리, 커피숍, 백화점 등에서 실제 환경잡음을 녹음하였다. 녹음된 잡음들은 다음 수식 (2)와 같이 깨끗한 음악 파일에 각각 다른 이득으로 더해져서 20dB, 15dB, 10dB, 5dB, 0dB의 SNR을 갖는 테스트 파일들로 구성되었다.

$$\widehat{N}(n) = \sqrt{\frac{P_s \times G}{P_N}} \times N(n) \quad (2)$$

$$\widehat{S}(n) = S(n) + \widehat{N}(n)$$

여기서 $S(n)$, P_s 는 음악 신호와 음악 신호의 파워를 나타내며, $N(n)$, P_N 은 실제 녹음된 잡음 신호와 파워, 그리고 G 는 이득을 맞추기 위한 파라메타이다.

모든 음악 파일들은 8,000Hz, 16bits, mono의 wave 파일로 변환하여 사용하였다.

2. ZCPA 특징의 성능 검증

그림 4는 ZCPA 특징의 성능을 검증하기 위한 것으로 기존의 연구에서 비교적 좋은 성능을 나타내는 것으로 알려진 필터뱅크 에너지(FBE)^[10]와 MFCC^[14]와의 성능을 비교하였다. 검색 성능은 다음 수식 (3)과 같이 비교하였다.

$$Recall (\%) = \frac{Retrieved Music}{Total Music} \times 100 \quad (3)$$

그림에서 보듯이 전반적으로 ZCPA 특징은 다른 특징들에 비해 잡음에 더 강한 성능을 가지고 있다. 거의 음 손실이 없는 20dB 환경에서는 ZCPA와 MFCC, FBE 특징 모두 유사한 성능으로 잘 동작하는 것을 알 수 있다. 그러나 잡음 영향이 심각해지는 15dB에서부터

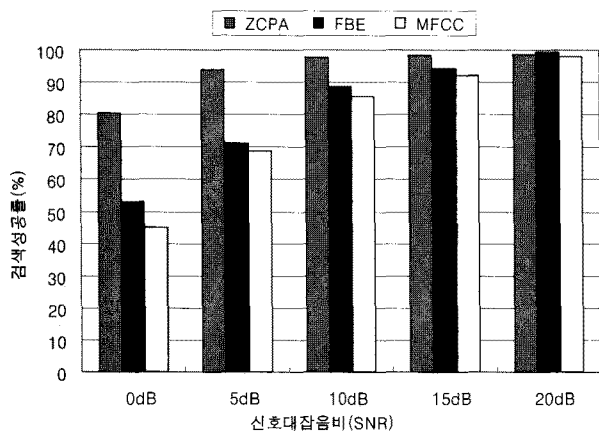


그림 4. 잡음 환경에서의 각 특징들의 성능 비교
Fig. 4. Compare the retrieval accuracy for extracted each feature in noise environment.

0dB 까지는 ZCPA와 다른 특징들 간에 약 20%~30%의 성능 차이가 있음을 확인할 수 있다. 특히 ZCPA 특징은 신호와 잡음의 에너지 비율이 동일한 0dB 상황에서도 약 80%의 검색 성공률을 나타내어 잡음에 매우 강한 특징을 가지고 있음을 알 수 있다.

3. 인덱싱 방법 및 검색 속도 비교

위의 그림 4에서의 결과는 ZCPA 특징의 잡음에 강한 특성을 확인하기 위하여 전수(Exhaustive) 검색 방법을 적용한 결과로써, 데이터베이스를 꼼꼼하게 검색하여 좋은 성능을 나타내기는 하지만 검색 속도를 고려한다면 실제 시스템에서는 적용하기 어려운 방법이다. 이에 본 논문에서는 ZCPA 특징의 성능을 보장하면서 검색 속도를 향상시킬 수 있는 방법을 연구하였다.

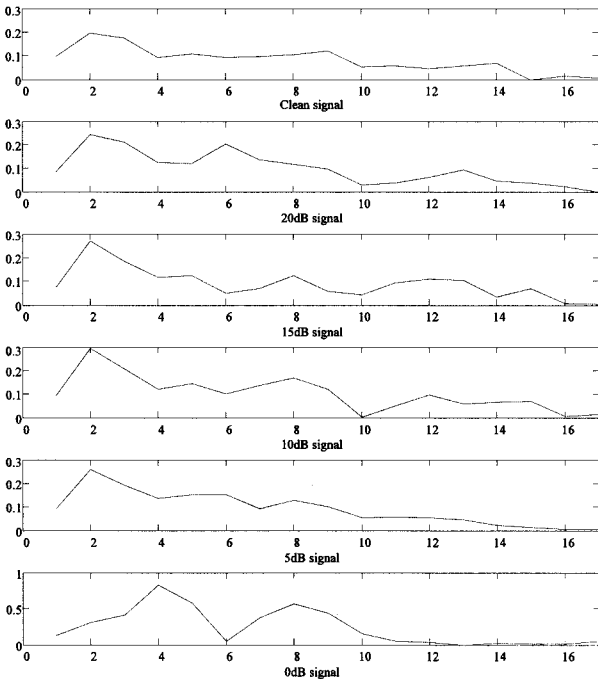
다음 그림 5는 환경 잡음에 의해서 왜곡되어 추출된 ZCPA 특징을 보상하여 인덱싱하는 방법을 보여주고 있다.

그림 5에서 x축은 주파수 히스토그램에서의 주파수 분할 구간의 개수로, 수식 (1)의 M에 해당한다. 본 논문에서는 주파수 히스토그램을 17개의 분할 구간으로 나누어 표현하였다. y축은 주파수 히스토그램의 크기이다. 그림 5(a)에서 보듯이 20dB에서부터 5dB까지의 신호에서 추출한 ZCPA 특징의 형태와 크기는 잡음으로 인해 Clean 신호와는 다른 왜곡된 모습을 가지게 된다. 그러나 그림 5(b)에서와 같이 수식 (4)의 MA(Moving Average) 필터링 을 거친 ZCPA 특징은 20dB ~ 0dB까지의 모든 잡음 환경에서 Clean 신호의 ZCPA 특징과 흡사한 형태를 보이고 있다.

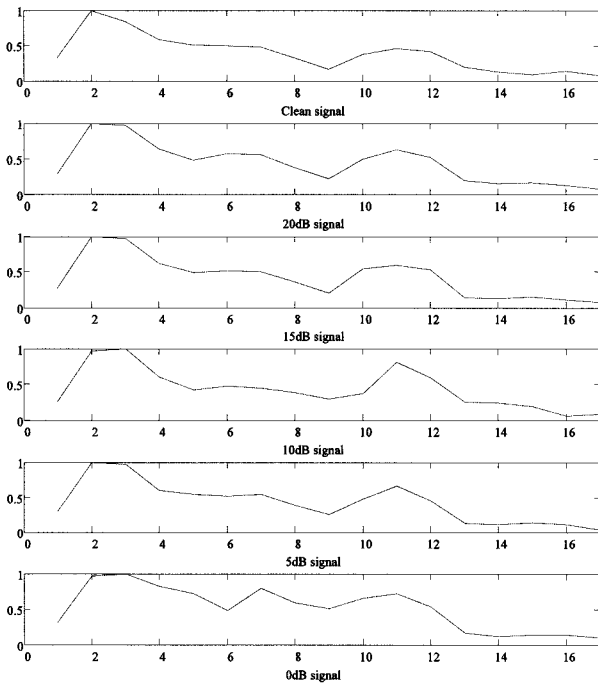
$$Z'(n, k) = \sum_{n=1}^{10} Z(n, k) \quad \text{단, } 1 \leq k \leq 17 \quad (4)$$

$$Z''(n, k) = Z'(n, k) / \max(Z'(n, k))$$

일반적인 MA 필터링 방법에서는 10 프레임씩 평균을 내면서 Z' 을 추출한 다음 프레임의 역수를 곱해주게 되지만, 본 시스템에서는 Z' 을 다시 Z' 의 최대값으로 나누어 정규화 시켰다. 본 작업은 5프레임씩 이동하며 1, 6, 11, ... 프레임에서 수행되며, 그림 5(b)에서 보듯이 MA 필터링 과정을 거친 ZCPA 특징은 잡음의 영향으로부터 비교적 자유롭기 때문에 검색 시스템의 인덱스로 사용하여도 손색이 없다. 본 논문에서는 검색 인덱스를 생성하기 위하여 그림 6과 같이 17개의 분할 구간으로 나누어있는 주파수 히스토그램을 0~7까지의 8구



(a) ZCPA



(b) MA 필터링 후의 ZCPA

그림 5. (a) 잡음 환경에서의 ZCPA

(b) MA필터링 후의 ZCPA

Fig. 5. (a) ZCPA feature for noise environment

(b) ZCPA feature after the MA filtering.

간으로 재분할 하였다. 재분할된 각 구간에서의 ZCPA 에너지를 구해 에너지가 큰 순서대로 정렬하여 검색을 위한 인덱스를 생성한다.

그림 7은 본 시스템의 검색 알고리즘을 설명한 것이

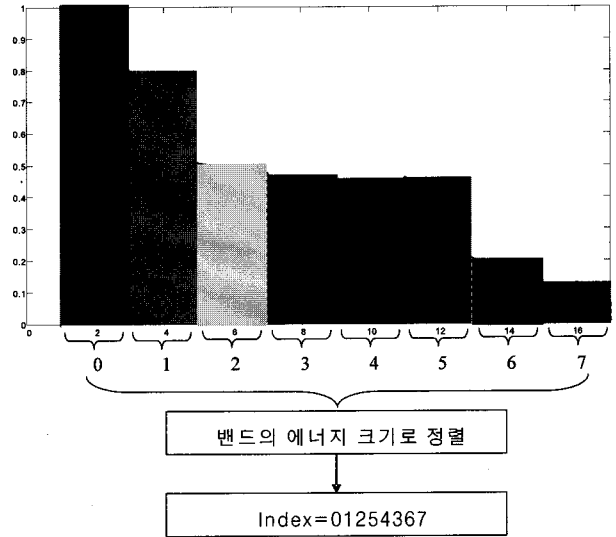


그림 6. 인덱스 생성 방법

Fig. 6. Method for creating index.

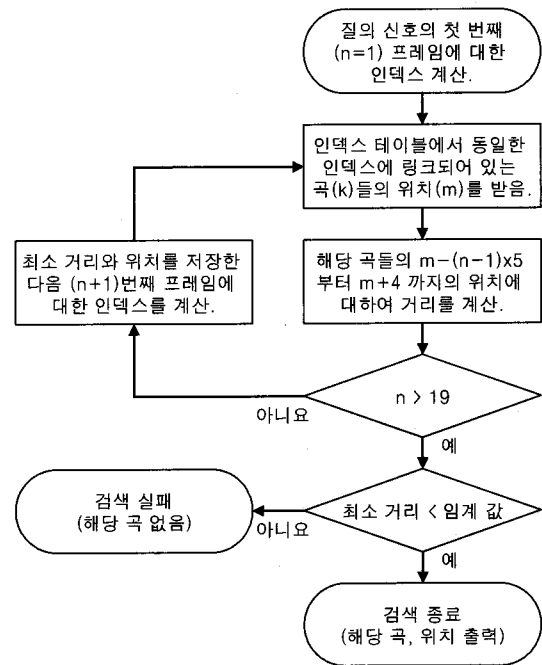


그림 7. 검색 알고리즘

Fig. 7. Proposed retrieval algorithm.

다. 제안 시스템은 1.5초 길이의 신호를 질의로 입력 받아 32ms를 한 프레임으로 50%씩 이동하면서 특징을 추출하게 된다^[17]. 그 결과 약 90 프레임에서 각 17개의 분할 구간으로 나누어 구해진 ZCPA 특징이 추출되고, 추출된 특징을 이용하여 5 프레임씩 이동하며 1, 6, 11, ... 프레임에서 인덱스를 생성하면 한 곡의 질의 당 총 19개의 인덱스가 생성된다.

먼저 1.5초 길이의 질의 음악이 입력되면 우선 인덱스 테이블에서 질의 신호의 첫 번째(n=1) 인덱스와 동일한 인덱스를 검색하여, 그 곳에 링크되어 있는 곡명

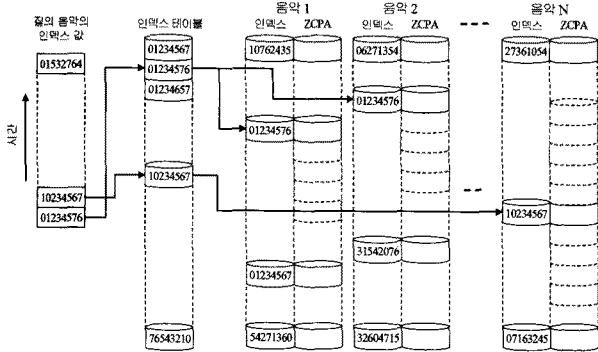


그림 8. 인덱스 검색의 예
Fig. 8. Example of indexed retrieval.

(ID)과 해당 위치(프레임)를 받아온다. 만약, 인덱스 테이블에 k 라는 곡의 m 번째 프레임이 연결되어 있다면 m 번째 프레임으로부터 $m+89$ 번째 프레임(1.5초)까지의 ZCPA 특징을 질의 음악의 ZCPA 특징과 비교하여 유사도 거리를 계산하며, 1 프레임씩 이동하면서 ($m+4$) 번째 프레임~($m+93$) 번째 프레임까지의 거리를 비교한다. 이는 인덱스 생성 시 5프레임씩 이동하면서 작업을 하였기 때문에 정확한 위치를 찾기 위한 방법이다. 이렇게 계산된 거리 값 중 최소 거리 값과 해당 위치를 저장한 후, 질의 음악의 두 번째($n=2$) 인덱스와 동일한 인덱스에 연결되어 있는 곡의 위치를 받아온다. 두 번째 인덱스부터는 m 번째 프레임~ $m+89$ 번째 프레임을 비교하지 않고, $m-(n-1)*5$ 번째 프레임~ $m-(n-1)*5+89$ 번째 프레임을 시작으로 하여 ($m+4$) 번째 프레임~($m+93$) 번째 프레임까지의 거리를 계산한다. 이러한 검색 방법은 1.5초의 질의 신호가 입력되는 동안 잡음이 아무리 심하게 더해진다 하여도, 적어도 입력 신호의 약 1/10의 길이에 해당하는 176ms 정도는 잡음으로부터 자유로울 수 있다는 가정에서 출발하였다(수식 (2) 참조). 만약, 질의 음악으로부터 생성된 19개의 인덱스 중 1개의 인덱스만이라도 잡음으로부터 자유롭다면 전수 검색(Exhaustive retrieval) 방법의 성능을 상당부분 보장할 수 있게 된다.

본 검색 방법에서 19 번째 인덱스와 동일한 인덱스에 링크되어 있는 곡들에 대해서는 각 곡당 $m-(19-1)*5 \sim m+4$ 까지 95번의 거리를 비교하게 되어 검색 속도에 영향을 주게 된다. 그러나 이러한 문제는 질의 신호의 특징 값은 변하지 않고, 비교하고자하는 DB 내의 신호의 특징 값만이 FIFO(First In First Out) 형태로 이동하면서 변화하기 때문에, 최초로 계산된 거리 값을 저장한 후 새로이 유입되는 프레임의 특징과 빠져나가는 프레임의 특징 간의 차이만을 계산하여 간단히 거리

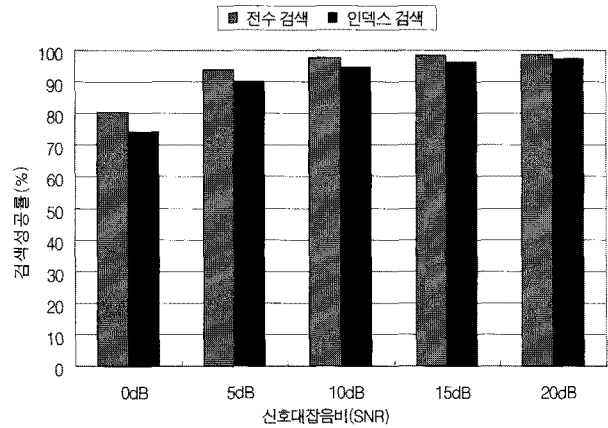


그림 9. 전수 검색과 인덱스 검색 방법 간의 검색 성공률 비교

Fig. 9. Compare the retrieval accuracy for the exhaustive retrieval and indexed retrieval.

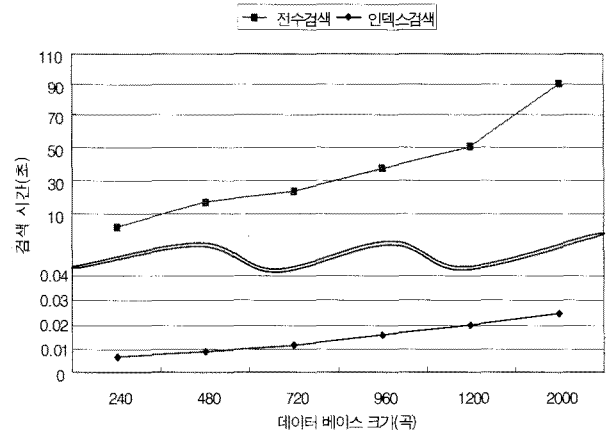


그림 10. DB의 증가에 따른 검색 시간

Fig. 10. Compare the retrieval time for the increasing of DB size.

값을 얻을 수 있기 때문에 큰 문제가 되지 않는다. 다음 그림 9는 전수 검색 방법의 검색 성공률과 본 논문에서 제안한 인덱스 검색 방법의 검색 성공률을 비교한 그래프이다. 모든 특징 값과 인덱스가 일대일 대응이 아닌 다대일 대응됨으로써 인덱스 검색시 약간의 성능 저하가 발생하였으나, 전수 검색 방법의 대응으로 무리가 없음을 확인할 수 있다.

그림 10은 본 논문에서 적용한 검색 방법이 실제 시스템에서 검색 시간을 얼마나 단축시켜 주는지에 대한 실험 결과이다. DB를 240곡에서 2000곡까지 단계별로 늘려가면서 전수 검색 방법과의 검색 속도를 비교 측정해보았다. 전수 검색의 경우 데이터베이스의 크기가 2 배 증가할 때 마다 곡당 평균 검색 시간이 약 2배 이상 증가하는 모습을 보이고 있으나, 인덱스 검색의 경우는 약간의 시간만이 더 소요되고 있다. 또한, 검색 속도의 향상률을 비교한다면 전수 검색 시보다 99% 이상 속도

가 향상되어, 실제 서비스를 위하여 10만곡 정도의 데이터베이스를 구축할 경우 약 2~4초 정도의 평균 검색 시간이 예상된다.

V. 결 론

본 논문에서는 모바일 환경에서 적용 가능한 잡음에 강인한 내용기반 음악 검색 시스템을 구축하였다. 제안된 시스템에서는 잡음에 강인한 ZCPA 특징을 사용하여 0dB의 상황에서도 약 80% 정도의 검색 성공률을 나타내었다. 또한, ZCPA 특징을 이용한 인덱스 검색 방법을 통하여 99% 가까이 검색 속도를 향상시킬 수 있었고, 10만곡 정도의 데이터베이스를 구축할 경우 약 2~4초 정도의 평균 검색 시간이 예상되어 실제 서비스를 위하여 활용 가능하다. 향후 연구로서는 잡음에 보다 강인한 ZCPA 특징을 개선 발전시키는 것과 인덱스 검색의 오류율을 줄이는 연구를 지속하고 있다.

참 고 문 헌

- [1] "2005 음악산업백서," 문화관광부.
- [2] E. Wold, T. Blum, D. Keislar and J. Wheaton, "Content-Based Classification, Search and Retrieval of Audio," *IEEE Multimedia*, Vol. 3, No. 3, pp.27-36, Fall 1996.
- [3] G. Tzanetakis, "Manipulation, Analysis and Retrieval Systems for Audio Signals," Princeton Computer Science Technical Report TR-651-02, June, 2002.
- [4] G. Li and A. Khokhar, "Content-based Indexing and Retrieval of Audio Data using Wavelets," *IEEE Int. Conf. on Multimedia and Expo*, 2000.
- [5] B. Logan, "Music recommendation from song sets," *International Symposium on Music Information Retrieval*, October, 2004.
- [6] Sheng Gao, Chin-Hui Lee and Qi Tian, "Indexing with Musical Events and Its Application to Content-Based Music Identification," *17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, UK, August 2004.
- [7] C. J. Burges, J. C. Platt and S. Jana, "Distortion Discriminant Analysis for Audio Fingerprinting," *IEEE Trans. on Speech and Audio Proc.*, Vol.11, No.3, pp.165-174, 2003.
- [8] E. Allamanche, etc. "Content-based identification of audio material using MPEG-7 low level description," *International Symposium on Music Information Retrieval*, October, 2001.
- [9] S. Sukittanon, L. Atlas, and J. Pitton, "Modulation-scale analysis for content identification," *IEEE Transactions on Signal Processing*, Volume 52, Issue 10, pp. 3023-3035, October, 2004.
- [10] J. Haitsma, and T. Kalker, "A Highly Robust Audio Fingerprinting System," *International Symposium on Music Information Retrieval*, October 2002.
- [11] Shazam, <http://www.shazam.com>.
- [12] Doh-Suk Kim, Soo-Young Lee, Rhee M. Kil, "Auditory Processing of Speech Signals for Robust Speech Recognition in Real-World Noisy Environments," *IEEE Transactions on Speech and Audio Processing*, Vol. 7, No. 1, Jan., 1999.
- [13] D. Fragoulis, G. Roussopoulos, Th. Panagopoulos, C. Alexiou, C. Papaodysseus, "On the automated recognition of seriously distorted musical recordings," *IEEE Transactions on Signal Processing*, Vol. 49, No 4, pp. 898-908, July 2000.
- [14] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling," *International Symposium on Music Information Retrieval*, October 2000.
- [15] E. Allamanche, J. Herre, O. Hellmuth, B. Bernhard Fröbach, and M. Cremer, "AudioID: Towards Content-Based Identification of Audio Material," *100th AES Convention*, Amsterdam, The Netherlands, May 2001.
- [16] O. Ghitza, "Auditory models and human performance in tasks related to speech coding and speech recognition," *IEEE Transactions on Speech and Audio Process*, Vol. 2, pp. 115-132, Jan., 1994.
- [17] Hyoung-Gook Kim, Ki-Wan Eom, "Robust Music Identification using Long-Term Dynamic Modulation Spectrum," *The Journal of the Acoustical Society of Korea*, Vol. 25, No. 2E, pp. 69-73, June, 2006.

저 자 소 개



윤 원 중(정회원)

2003년 상명대학교 정보통신학과
학사 졸업.

2005년 단국대학교 컴퓨터과학 및
통계학과 석사 졸업.

현재 단국대학교 컴퓨터과학 및
통계학과 박사 과정.

<주관심분야 : 음성 및 음향신호처리, 멀티미디어
신호처리, DSP 시스템 구현>



박 규 식(정회원)

1986년 Polytechnic University
전자공학과 학사 졸업.

1988년 Polytechnic University
전자공학과 석사 졸업.

1993년 Polytechnic University
전자공학과 박사 졸업.

1994년~1996년 삼성전자 마이크로사업부, 선임
연구원.

1996년~2001년 상명대학교 컴퓨터·정보통신
공학부 조교수.

2001년~현재 단국대학교 컴퓨터학부 교수.

<주관심분야 : 음성 및 음향신호처리, 멀티미디어
신호처리, DSP 시스템 구현>