

## 내용 기반 음악 검색의 문제점 해결을 위한 전처리

정 명 범\*, 성 보 경\*\*, 고 일 주\*\*\*

# Pretreatment For The Problem Solution Of Contents-Based Music Retrieval

Myoung-Beom Chung\*, Bo-Kyung Sung\*\*, Il-Ju Ko\*\*\*

### 요 약

본 논문에서는 오디오를 내용기반으로 분석, 분류, 검색하기 위하여 사용되어 온 특징 추출 기법의 문제점을 제시하며, 새로운 검색 방법을 위해 하나의 전처리 과정을 제안한다. 기존 오디오 데이터 분석은 샘플링을 어떻게 하느냐에 따라 특징 값이 달라지기 때문에 같은 음악이라도 다른 음악으로 인식될 수 있는 문제를 갖고 있다. 따라서 본 논문에서는 다양한 포맷의 오디오 데이터를 내용 기반으로 검색하기 위해 PCM 데이터의 파형 정보 추출 방법을 제안한다. 이 방법을 이용하여 다양한 포맷으로 샘플링 된 오디오 데이터들이 같은 데이터임을 발견 할 수 있으며, 이는 내용기반 음악 검색에 적용 할 수 있을 것이다. 이 방법의 유효성을 증명하기 위해 STFT를 이용한 특징 추출과 PCM 데이터의 파형 정보를 이용한 추출 실험을 하였으며, 그 결과 PCM 데이터의 파형 정보 추출 방법이 효과적임을 보였다.

### Abstract

This paper presents the problem of the feature extraction techniques that has been used a content-based analysis, classification and retrieval in audio data and proposes a course of the preprocessing for a new contents-based retrieval methods. Because the feature vector according to sampling value changes, the existing audio data analysis is problem that same music is appraised by other music. Therefore, we propose waveform information extraction method of PCM data for retrieval audio data of various format to contents-based. If this method is used, we can find that audio datas that get into sampling in various format are same data. And it may be applied in contents-based music retrieval system. To verity the performance of the method, an experiment was done feature extraction using STFT and waveform information extraction using PCM data. As a result, we could know that the method to propose is effective more.

▶ Keyword : 오디오 특징 벡터 추출(Audio feature vector extraction), 오디오 정보 검색(Audio Information Retrieval), 내용 기반 검색(Contents-based retrieval)

• 제1저자 : 정명범    • 교신저자 : 고일주

• 접수일 : 2007.10.29, 심사일 : 2007.11.6, 심사완료일 : 2007.11.20.

\* 송실대학교 IT대학 미디어학과 박사과정    \*\* 송실대학교 IT대학 미디어학과 박사과정

\*\*\* 송실대학교 IT대학 미디어학과 조교수

※ 본 연구는 송실대학교 교내연구비 지원으로 이루어졌음

## 1. 서론

최근 컴퓨터 환경의 발달과 폭 넓은 인터넷의 확산으로 인하여 사용자들은 이미지, 오디오, 비디오 등 다양한 멀티미디어 데이터를 제공받게 되었다. 또한 다양한 소프트웨어의 제공으로 사용자들은 데이터를 받기만 하는데 그치지 않고, 받은 데이터를 가공, 재생산을 하여 다시 다른 사람들에게 제공하는 중간자 역할을 하기도 한다. 따라서 같은 유형의 데이터도 사용자들의 가공에 의해 비슷한 형태의 데이터로 만들어진다.

오디오 데이터의 경우에서도 이와 유사한 흐름이 나타난다. 초기 음반 시장은 LP와 Tape을 시작으로 하여, 1980년대에 오디오 CD가 개발되어 보급 되었다. 이때까지는 컴퓨터 성능뿐 아니라 다른 포맷의 변환 장치가 없어 사용자는 제공받는 데이터만을 사용하였다. 그러나 1987년 독일의 IIS(Institut für Integrierte Schaltungen)라는 곳에서 심리음성학에 의거한 인지 압축 기술에 의하여 만들어진 MPEG Layer-3(MP3)가 보급되고, 그 이후 저장 장치와 컴퓨터 성능의 발달에 의해 WMA, OGG, ASF와 같은 다양한 포맷을 갖는 오디오 데이터들이 나타났다. 이러한 다양한 포맷은 사용자들이 직접 오디오 데이터를 가공하여 자신에게 맞는 데이터로 변형을 시킬 수 있게 하였으며, 그에 따라 하나의 음악은 같은 노래임에도 불구하고 다양한 포맷, 크기를 갖게 되었고, 현재 사용되고 있는 전체 오디오 데이터의 종류와 양은 셀 수 없을 정도로 방대해졌다.

오디오 데이터의 증가는 사용자들의 검색 요구를 증가시켰으며, 이를 위한 검색 시스템에 대한 연구 역시 활발히 진행되었다. 초기의 검색은 수작업을 통해 각각의 제목이나, 노래 가사, 가수 등을 입력한 텍스트 기반의 분류 시스템에 저장하여 검색을 제공하였다. 즉, 데이터가 있다하더라도 텍스트 내용이 맞게 입력되지 않는다면 검색되지 않는 문제가 있었다. 이를 보완하기 위한 방법으로 기존의 수작업이 아닌, 정보의 내용을 수학적으로 분석하여 구조화된 기준에 따라 대표적인 특징을 추출하고, 그러한 특징을 토대로 데이터를 체계적인 구조로 색인화 하는 내용 기반의 정보 검색 시스템이 제공되었다. 내용 기반 특징 벡터 추출 연구는 허밍(Humming)과 같은 음악의 한 부분을 이용하여 전체를 찾는 음악 매칭 검색 [1-3], 비슷한 음악적 특징 벡터를 갖는 데이터로부터 음악 장르를 분류하는 장르 분류 알고리즘 [4-6], 음악의 오디오 신호를 이용하여 음악의 분위기를 알아내는 음악 분위기 묘사 [7, 8] 등이 있으며, 이러한 연구는 다양한 분야에 적용되어 사용되고 있다.

그러나 오디오 데이터를 내용 기반의 정보 검색 시스템에 적용하기 위해서는 데이터가 고유하지 않다는 문제점이 있다. 기존의 내용 기반 정보 검색을 위해서는 오디오 데이터의 특징 벡터 추출이 필요하다. 특징 벡터 추출 방법은 전처리 과정을 통해 오디오의 PCM 데이터 값을 얻고, 그 값으로부터 파형 정보를 얻어 FFT(Fast Fourier Transform)나 STFT(Short Time Fourier Transform) 변환을 시킨다. 여기서 Spectral Centroid, Rolloff, Flux 등의 특징 벡터 값을 얻을 수 있으며 이 값을 내용 기반 정보 검색에 사용한다. [9] 즉, 같은 값으로 샘플링 된 오디오 데이터는 내용 기반 정보 검색에 의해 정확히 일치하는 특징 벡터를 얻을 수 있다. 그러나 이러한 방법은 앞서 언급한 오디오 데이터의 다양한 포맷 때문에 샘플링 된 값에 따라 같은 노래지만 다른 노래로 인식되는 문제점이 발생한다.

본 논문에서는 이러한 문제점을 해결하기 위한 방법으로 PCM 데이터의 파형 정보 추출법을 제안한다. 기존의 내용기반 분석에서는 파형 정보를 FFT나 STFT를 이용하여 주파수 영역으로 변환 후 내용 기반 분석을 하였다. 그러나 본 논문에서는 PCM 데이터의 파형 정보로부터 일정 구간 내에 최대 파형의 90% 이상 집중되어 있는 위치를 찾는 방법으로서, 기존의 내용기반 분석과는 달리 파형 자체를 다루어 특징 위치를 추출한다. PCM 데이터의 파형 정보 추출법은 일정 구간 내에 최대 파형의 90% 이상이 집중되어 있는 위치를 찾는 것이다. 이 방법은 다양한 포맷으로 샘플링 된 오디오 데이터라 할지라도 파형의 형태가 변하지 않는 특성을 이용한 것으로 내용 기반 정보 검색에 효과적인 결과를 얻을 수 있다.

제안 방법의 유효성을 검증하기 위해 같은 음악의 다양한 포맷에 오디오 데이터를 실험 데이터로 하였으며, STFT 특징 벡터 추출과 PCM 데이터 파형 정보 추출을 하여 그 값을 비교하였다. 이때 STFT의 특징 벡터는 Spectral Centroid, Rooloff, Flux를 이용하였다.

논문의 구성은 다음과 같다. 2장에서 내용기반 분류 및 검색 연구에 대한 관련 연구를 언급하고, 3장에서 내용기반 검색 방법의 문제점을 설명한다. 4장에서 제안 방법을 구체적으로 설명 하고, 5장에서 제안한 방법의 유효성을 검증하기 위해 비교 실험과 분석을 하며, 6장에서는 결론을 제시한다.

## II. 관련 연구

기존의 내용기반 오디오 장르 분류 및 검색에 관련된 연구는 1997년 미국 Audible Magic 사의 Muscle-Fish [10]가 그 효시로서, 크게 DSP(Digital Signal Processing) 기술

을 이용한 방법과 MIDI 파일의 음악 표기 정보를 이용하는 2 가지 방법이 있다. DSP를 이용한 방법은 모든 오디오 파일 포맷에 적용이 가능하며, 오디오의 Pitch, Timber, Harmony 등의 특징을 추출하여 장르 분류, 검색을 하는 것이다.

DSP 기법의 논문[10]에서는 여러 가지 15초미만의 음향 효과들에 대하여 신호의 크기, Pitch, 밝음, 대역폭, 하모니 등의 특징을 추출하였다. 그러한 음향 효과들의 구분으로부터 점차 연구는 오디오 데이터의 장르 분류를 하기 시작하였다. 특히 논문 [11]에서 G. Tzaneta- kis는 음악 장르의 자동 분류를 위해 STFT 기반의 오디오 특성과 Wavelet Transform 기반의 Rhythm, Pitch등의 오디오 특성을 추출하여 Blues, Classical, Country, Disco, Hiphop, Jazz등 10가지의 음악 장르 분류에서 약 70% 정도의 성공률을 보였으며, 분류기로는 k-NN, Gaussian, GMM 등을 사용하였다. 논문 [12]에서는 앞서 언급한 논문과 달리 Wavelet Basics에 의한 DWCH라는 특징들을 추출하여 SVM, GMM, LDA, KNN 등 여러 가지 분류기로 실험을 하였고, 그 논문에서 사용한 FFT, MFCC가 장르 분류에 많은 영향을 미치며, Beat와 Pitch는 큰 영향을 미치지 않음을 알 수 있다.

또한 논문 [13]에서는 여러 가지 특징들의 추출로부터 SFS를 이용한 특징 벡터를 최적화를 제안하였다. 기존 특징 벡터를 모두 추출한 다음(54차), SFS (Sequential Forward Selection) 기법을 이용하여 1/5에 해당하는 10 차의 최적 특징 벡터만을 선정하여 오디오 장르 분류에 사용하였다.

이와 같이 앞선 연구 논문에서 나타나는 공통적인 오디오 데이터의 특징 벡터는 STFT 기반의 Spectral Centroid, Rolloff, Flux등이며, MFCC[14] 기반의 인간의 청각 특성에 맞는 Mel- Frequency 스케일로 변환한 13차 계수 특징 벡터라는 것을 알 수 있다.

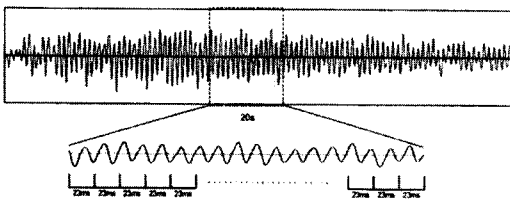


그림 1. STFT를 이용한 특징 벡터의 추출  
Fig.1 Extraction of feature vector using STFT

내용 기반의 검색 시스템에서 주로 사용되는 특징 벡터 중 STFT 기반의 특징 계수인 Spectral Centroid, Rolloff,

Flux를 얻는 방법은 다음과 같다. 입력 받은 오디오 데이터를 22050Hz, mono로 전처리하며, 20초 분량의 PCM 데이터를 (그림 1)과 같이 23ms 크기의 Hamming window를 중복되지 않게 이동하면서 각 23ms 프레임으로부터 STFT를 적용하여 값을 구한다. 각각의 특징 계수를 구하는 방법은 아래의 식과 같다.

$$C_t = \frac{\sum_{n=1}^N (M_t[n] \times n)}{\sum_{n=1}^N M_t[n]} \dots\dots\dots (식 1)$$

Spectral Centroid(식 1)는 STFT의 크기 스펙트럼의 중심을 뜻하며, 스펙트럼의 형태를 측정하는 방법 중의 하나이다. 여기서  $M_t[n]$ 은 프레임  $t$ 와 주파수 Bin (Frequency Bin)  $n$ 에서의 스펙트럼 크기에 해당한다.  $C_t$ 는 프레임  $t$ 의 Centroid 값을 나타내며,  $N$ 은 프레임  $t$ 내에 있는 주파수Bin의 최대값이다. Centroid의 값이 높을수록 더 높은 주파수에서 선명한 음질을 나타낸다.

$$\sum_{n=1}^R M_t[n] = 0.8 \times \sum_{n=1}^N m_t[n] \dots\dots\dots (식 2)$$

Spectral Rolloff(식 2)는 크기 스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지를 보여준다. 아래의 식은 스펙트럼 분포의 80%가 집중되어 있는 주파수  $R_t$ 을 구하기 위한 것이다.  $R_t$ 는 크기 스펙트럼 전체를 더하여 80% 값을 얻은 후 다시 크기 값을 1에서부터  $R_t$ 까지 합으로 비교하여 구할 수 있으며, 이때의  $R_t$ 는 크기 분포의 80%가 집중되어 있는 주파수이다.

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2 \dots\dots\dots (식 3)$$

Spectral Flux(식 3)는 연속된 스펙트럼의 분포에서 스펙트럼의 변화의 양을 측정하는 것이다. 아래의 식은 연속된 스펙트럼 분포에서 정규화 된 크기들 간의 차이를 제곱해서 Spectral Flux를 구하기 위한 것이다.  $N_t[n]$ 과  $N_{t-1}[n]$ 은 주파수 Bin  $n$ 과  $n-1$ 에 해당하는 크기 값을 정규화 한 것이며, 이로부터 스펙트럼의 변화량인  $F_t$ 를 구한다.

### III. 내용 기반 검색 방법의 문제점

사람들이 들을 수 있는 소리의 영역에는 한계가 있다. 구체적으로 20Hz~22,000Hz 정도의 주파수 대역의 소리를 귀로 감지 할 수 있는데, 이러한 아날로그 원음을 완벽히 디지털 형태로 재생하려면 원음이 가지는 주파수의 2배 이상이 되는 주파수로 샘플링을 해야 한다. 따라서 오디오 시디는 1초에 44,100번 소리를 따내는데 이 때 한번 따낸 소리를 16Bit(2Byte)로 표현하기 때문에 1분이라는 시간동안  $60s \times 44,100 \times 2\text{byte}$ 를 계산하면 대략 5MB 정도가 나오게 된다. 여기에 스테레오로 하면 대략 10MB 정도가 나온다.

MP3는 이러한 파일을 압축하는 파일 포맷이다. 압축의 원리는 사람의 귀로 들을 수 없거나 듣지 않아도 되는 부분은 버리고 디지털화함으로써 데이터를 줄이는 것이다. 아날로그 음악을 디지털화하는 과정에서 사람이 들을 수 있는 가청 주파수 범위를 넘거나, 특정 악기 소리 뒤에 붙는 여운 등을 뺏으므로 파일의 용량을 줄일 수 있다. 이때 PCM 데이터의 값은

FFT를 거쳐 주파수 영역으로 바꾸어 계산하는데 샘플링 수가 클수록 원본 데이터와 가깝게 만들 수 있으며, 샘플링 수가 작을수록 음질은 떨어지지만 데이터 용량을 작게 만들 수 있다. [그림 2]는 톱니파형을 샘플링 할 때 FFT를 이용하여 파형을 주파수 영역으로 변환하고 64k, 128k로 샘플링 한 것이다.

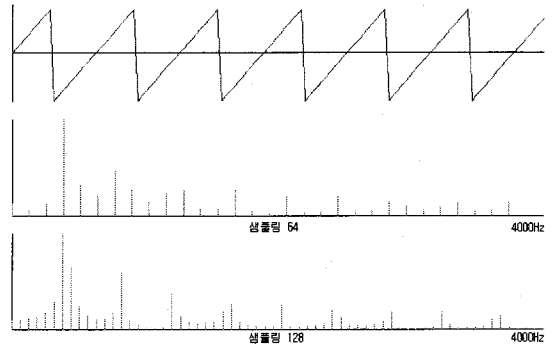


그림 2. FFT를 이용한 톱니파형의 샘플링  
Fig.2 Sampling of sawtooth waveform using FFT

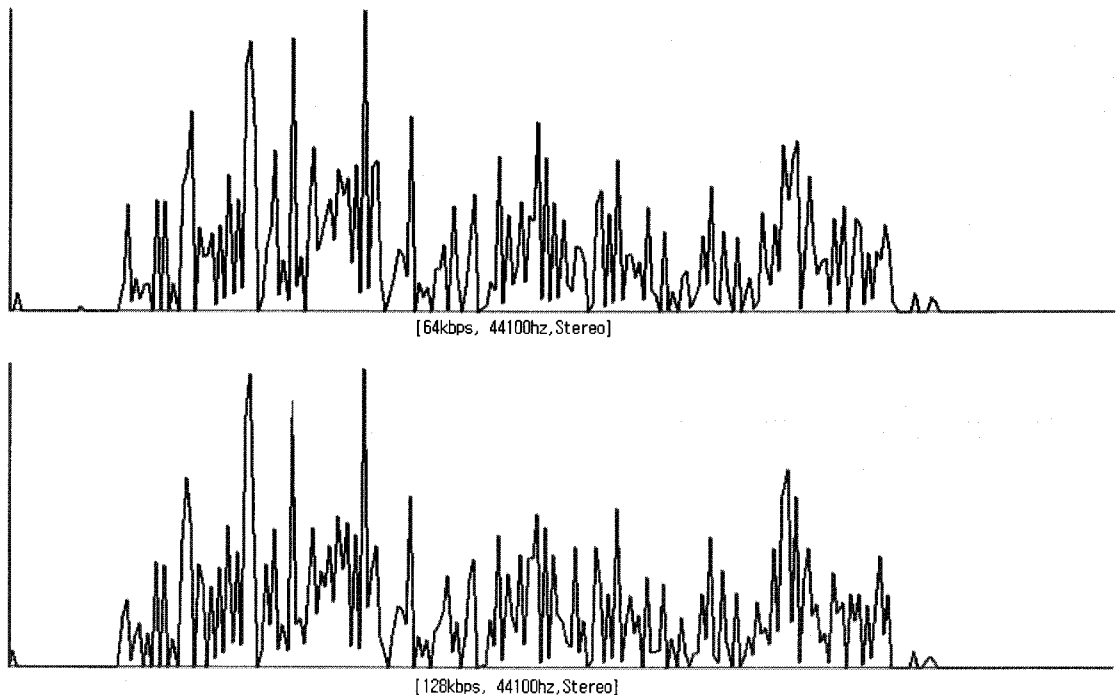


그림 3. 동일 오디오 데이터를 64kbps, 128kbps로 샘플링 후 구한 PCM 데이터 파형  
Fig.3 PCM data waveform of 64kbps, 128kbps sampling rate using same audio data

[그림 2]에서 파형을 주파수 영역으로 변환한 것에서 보면 64kbps로 샘플링 한 것보다 128kbps로 샘플링 한 것이 더 조밀하고, 64kbps에서 없는 주파수 값들이 존재한다. 즉, 샘플링 값에 따라 MP3로 압축 변환이 될 때 데이터 값이 달라짐을 의미한다. 데이터 값의 달라짐은 MP3에서 뿐만 아니라 Ogg, Vbr, WMA에서도 변화가 된다.

현재의 내용 기반 검색 시스템에서는 이러한 압축에 따른 데이터의 변형에 관심을 갖지 않고, 기존에 가지고 있는 wav 데이터만을 기반으로 하기 때문에 문제가 생기지 않는다. 하지만 실제 서비스, 사용되고 있는 오디오 데이터는 같은 음악이지만 다양한 포맷을 갖고 있어 내용 기반 특징 검색 시스템의 특징 벡터 추출을 이용하면 다른 음악으로 인식 될 수밖에 없다. 따라서 실제 사용될 수 있는 내용 기반 검색을 위한 다른 방법이 요구 된다.

#### IV. PCM 데이터의 파형 정보를 이용한 내용 기반 검색

PCM 데이터의 파형 정보 추출을 이용한 내용 기반 검색

은 기존의 내용 기반 검색에서 사용하듯이 PCM 데이터를 FFT를 이용하여 주파수 값으로 변환을 시키는 것이 아니라, PCM 데이터의 파형 형태 자체를 이용하는 것이다. [그림 3]은 '휘성 - 안되나요'라는 곡을 64kbps와 128kbps로 샘플링 한 MP3 파일의 PCM 데이터 파형이다.

[그림 3]에서 64kbps로 샘플링 한 PCM 데이터 파형과 128kbps로 샘플링 한 파형의 형태가 유사함을 알 수 있다. 따라서 이 파형들의 형태가 같음을 찾아내어 그 값을 특징 벡터로 얻는다면 이 특징 벡터를 내용 기반 검색 시스템에 사용할 수 있다.

파형의 형태 각각을 그대로 비교하는 DTW(Dynamic Time Warping)와 같은 연구가 예전부터 이루어져 왔으나, 이 방법은 짧은 구간의 음성 인식에 주로 이용이 가능한데 비해 오디오 데이터와 같이 많은 양의 정보를 가진 데이터간의 비교에는 상당한 무리가 따른다.

따라서 본 논문에서는 PCM 데이터의 값 중 최대의 값을 찾아낸 후 그 값의 90%에 해당하는 값을 구하고 그 값을 임계값으로 한다. 그 이후 일정 구간의 PCM 데이터 간격 내에 임계값 이상의 값을 갖는 데이터 개수를 구하고 이때 데이터의 개수가 가장 많은 곳을 찾는다. 이 부분이 PCM 데이터의

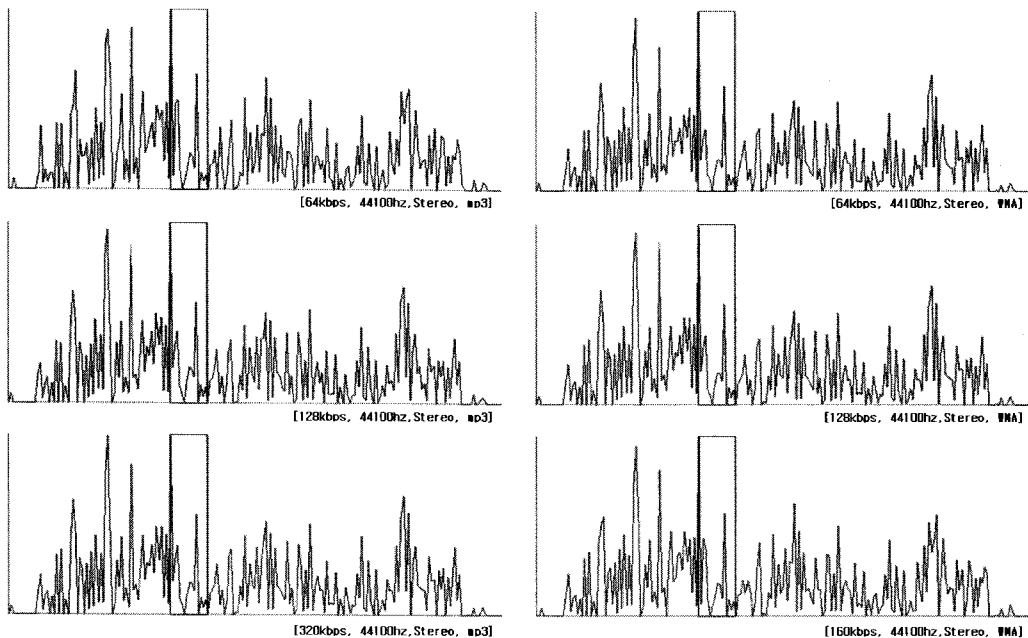


그림 4. 64, 128, 320kbps-MP3와 64, 128, 160kbps-WMA 각각의 파형 집중 위치 값  
Fig.4 Waveform concentration position of 64, 128, 320kbps-mp3 and 64, 128, 160kbps-wma

파형 중에서 높은 값이 가장 많이 집중되어 있는 부분이고, 이 위치가 여기서 제안하는 파형 집중 위치 값이다.

[그림 4]는 '취성 - 안되나요'라는 곡을 64kbps, 128kbps, 320kbps로 샘플링 한 MP3 파일과 64kbps, 128kbps, 160kbps로 샘플링 한 WMA 파일의 PCM 데이터 파형을 비교한 것이다. [그림 4]의 결과와 같이 64kbps, 128kbps, 320kbps로 변형했던 MP3 파일과 64kbps, 128kbps, 160kbps로 변형했던 WMA 파일이 모두 동일한 위치에서 파형 집중 위치 값이 나타남을 확인할 수 있다. 따라서 본 논문에서 제안하는 파형 집중 위치 값을 내용 기반 검색 시스템에 사용하면 다양한 포맷의 오디오 데이터에서도 같은 오디오 데이터를 검색해 낼 수 있을 것이다.

### V. 실험 및 분석

실험은 동일한 노래를 다양한 포맷으로 변형한 것을 오디오 데이터로 사용하였다. 제안 방법의 유효성을 검증하기 위해 먼저 기존 내용 기반 검색 시스템에서 사용하는 STFT의 기반의 특징벡터 Spectral Centroid, Rolloff, Flux에 대한 6차 특징 벡터를 추출하여 보았다. <표 1>은 '바이브 - Promise U'이라는 곡에서 6차 특징 벡터를 추출한 것이다.

표 1. 동일 음악의 다양한 포맷을 가진 데이터의 STFT 특징 벡터 값  
Table 1. STFT feature vector of various formats at same audio data

구분 (kbps)	Centroid		Rolloff		Flux		
	평균	분산	평균	분산	평균	분산	
MP3	64	118.10	150.09	243.39	76.80	0.19	3.05
	128	119.94	119.28	236.37	105.67	0.30	4.30
	320	120.03	118.06	235.97	119.72	0.21	3.19
WMA	64	120.05	117.76	236.59	101.31	0.44	8.35
	128	120.09	116.50	235.89	115.35	0.39	6.94
	160	120.08	116.65	235.92	115.31	0.39	7.06

<표 1>에서처럼 MP3와 WMA에서의 각 오디오 데이터의 특징 값이 샘플링 한 값에 따라 차이가 남을 볼 수 있다. 동일한 데이터를 가지고 제안한 방법인 PCM 데이터의 파형 집중 위치 값을 구하는 실험을 하였다. (이때 PCM 데이터 간격은 80으로 하였다.)

표 2. 동일 음악의 다양한 포맷을 가진 데이터의 시작 위치 값과 집중도  
Table 2. Start point and concentration position of various formats at same audio data

구분(kbps)	시작 위치	끝 위치	시작 위치값	집중도	
MP3	64	3484320	4139680	42	504
	128	3484320	4139680	42	516
	320	3484320	4139680	42	503
WMA	64	3484320	4139680	42	520
	128	3484320	4139680	42	449
	160	3484320	4139680	42	488

<표 2>에서 실험 결과를 보면 모든 포맷에서 집중도 값은 다르지만 시작 위치 값이 동일함을 볼 수 있다. 실험이 이 노래에 대한 우연한 결과가 아님을 보이기 위해 다음 실험으로 100곡의 오디오 데이터로 PCM 데이터의 파형 집중 위치 값 측정을 하였다.

<표 3>는 실험 결과의 일부를 가지고 온 것이며, 100곡 중 81곡이 같은 위치 값을 나타냄을 확인할 수 있다. 즉 다양한 포맷이더라도 동일한 곡이라면 제안한 방법으로 같은 위치를 추출할 수 있다. <표 3>의 오류가 나타난 두 곡 중 '윈타임-패 지나 청청'이라는 곡은 MP3의 64, 128, 320과 WMA의 128, 160에서는 같은 위치 값을 나타냈으나, WMA의 64kbps에서만 다른 위치를 나타냈다. 이 곡에서는 파형 집중 위치 값의 집중도가 다른 곡에 비해 200 이하인 작은 값을 나타낸다. 즉, 오류가 나타난 원인은 다른 위치에 비해 높은 값을 갖지 않아 그와 유사한 다른 위치의 값을 찾은 것으로 보인다.

이러한 오류는 모든 포맷 중 한 곳에서 나타나는 것이기 때문에 예외 처리 기법을 이용하여 그 포맷을 제외한 방법을 이용할 수 있을 것이다. 또 다른 오류를 나타내는 곡은 '상상 밴드 - 가지마가지마'라는 곡이다. 이 곡에서는 처음 구간 간격을 80으로 하였을 때 MP3의 64, 128, 320과 WMA의 160만 같은 위치를 찾았다. 그 밖에 WMA의 64와 128의 위치는 다른 곳을 나타냈다. 그러나 이를 보완하고자 2차적으로 구간 간격을 160으로 늘이고 계산을 하였을 때 모두가 같은 위치를 나타내었다. 즉 1차 실험인 구간 간격 80을 하였을 때 얻지 못한 값을 2차 실험 구간 간격 160으로 하였을 때 파형 집중 위치를 정확히 찾아낼 수 있었다.

결론적으로 제안된 방법은 실험으로부터 81%의 성공률을 얻을 수 있었으며, 이로부터 내용 기반 검색의 전처리 요소로 사용할 수 있는 가능성을 보였다.

표 3. PCM 데이터의 파형 집중 위치 값 측정  
Table 3. Measured waveform concentration position of PCM data

노래 제목	시작위치	위치값	MP3			WMA		
			64	128	320	64	128	160
브라운 아이즈 - 벌써 일년	4672160	71	331	419	493	485	495	492
노블레스 - 얼마나 행복하려고	2852960	51	207	190	209	222	195	210
궁OST - Perhaps Love(사랑인가요)	2583200	20	246	257	263	264	261	260
원타임 - 퀘지나 칭칭	6433440	104	169	150	174	(259)	172	186
김성재 - 말하자면	2828960	50	320	357	380	453	478	496
거북이 - 왜이래	3882960	63	258	293	307	385	391	393
박지원 - 느낌만으로	4622180	69	303	342	357	349	358	355
인디고 - 여름아 부탁해	2500720	16	227	276	290	284	297	304
노이즈 - 성형미인	4773150	73	263	269	280	277	296	284
상상밴드 - 가지마가지마 (간격160)	5614240	94	539	566	563	570	557	590

## VI. 결론 및 향후 연구 방향

오디오 데이터는 시간이 갈수록 더욱 많은 양의 데이터를 가지게 되며, 그 포맷 또한 더욱 다양해지고 있다. 이에 따라 오디오 데이터의 다양한 포맷에서도 같은 음악을 찾을 수 있는 내용 기반 검색 시스템의 개발이 필요하다. 여기서는 PCM 데이터의 파형 정보로부터 파형 집중 위치 값을 찾으므로 여러 가지 포맷을 갖는 음악이지만 동일한 음악인 경우 같은 위치를 얻을 수 있도록 하였다. 이 방법은 기존 내용 기반 검색에서의 주파수 영역으로 변환하는 방법과 차별이 되는 방법으로, 파형 정보 자체를 다룬다. 기존의 음악 분석에서 파형 정보를 다루지 않았던 이유 중 하나는 많은 데이터의 양을 일일이 비교하기 위해 엄청난 시간이 걸리기 때문이었다. 하지만 본 논문에서 제안하는 방법은 파형 정보의 특징 부분만을 추출하여 비교하기 때문에 기존의 문제점을 해결 할 수 있으며, 내용 기반 음악 분석에 새로운 접근이 될 것이다.

제안한 방법을 이용한 실험에서는 81%의 성공률을 나타내어 다양한 포맷에서 같은 음악을 찾을 수 있음을 볼 수 있었다. 따라서 내용 기반 검색 시스템에 사용할 수 있는 하나의 전처리 과정으로 적용 할 수 있을 것이다. 또한 여기서의 오류가 나타나는 곡의 경우 PCM 데이터의 간격을 변경하였을 때 다시 같은 위치를 찾아냄을 보였다. 즉, PCM 데이터의 간격을 임의로 정하는 것이 아니라 그 곡이 나타내는 비트(Beat), 템포(Tempo)를 추출하여 적용하면 보다 나은 결과

를 얻을 수 있을 것으로 예상된다.

앞으로의 연구는 오류로 나타났던 문제점을 해결하기 위해 오디오 데이터로부터 템포를 추출하여 그 값을 논문에서 제안한 방법으로 적용할 것이다. 이미 오디오 데이터의 템포를 추출하는 연구는 많이 이루어져 왔으므로 그 연구 내용을 검토하고 실험하여 제안한 방법에 맞는 템포를 추출 할 수 있을 것이다. 이를 기반으로 내용 기반 검색 시스템에 보다 나은 검색을 위한 특징 벡터를 모색하는 것이 향후 연구 과제이다.

## 참고문헌

- [1] T. Sonoda, M. Goto and Y. Muraoka, "A WWW-based melody retrieval system," Proc. ICMC 1998, pp. 349 - 352, 1998.
- [2] S. Pauws, "CubyHum: A fully operational query by humming system," Proc. ISMIR 2002, pp. 187 - 196, 2002.
- [3] T. Sonoda, T. Ikenaga, K. Shimizu and Y. Muraoka, "The design method of a melody retrieval system on parallelized computers," Proc. WEDELMUSIC 2002, pp. 66 - 73, 2002.
- [4] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech Audio Proc., 10, pp. 293 - 302, 2002.

- [5] S. Dixon, E. Pampalk and G. Widmer, "Classification of dance music by periodicity patterns," Proc. ISMIR 2003, pp. 159 - 165, 2003.
- [6] M. F. McKinney and J. Breebaart, "Features for audio and music classification," Proc. ISMIR 2003, pp. 151 - 158, 2003.
- [7] K. Kashino and H. Murase, "A sound source identification system for ensemble music based on template adaptation and music stream extraction," Speech Commun., 27, pp. 337 - 349, 1999.
- [8] J. Eggink and G. J. Brown, "A missing feature approach to instrument recognition in polyphonic music," Proc. ICASSP 2003, pp. V-53 - 56, 2003.
- [9] J. M. Gray. "An Exploration of Musical Timbre," PhD thesis, Dept. of Psychology, Stanford University, 1975.
- [10] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," IEEE Multimedia, 3(2), pp. 27-36, 1996.
- [11] George Tzanetakis and Perry Cook, "Musical genre classification of audio signals," IEEE Transaction on Speech and Audio Processing, no.5, pp. 293-302, 2002.
- [12] Tao Li, Mitsunori Ogihara, and Qi Li, "A comparative study on content-based music genre classification," SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp. 282-289, 2003.
- [13] 윤원중, 이강규, 박규식, "내용기반 오디오 장르 분류를 위한 신호 처리 연구," 전자공학회논문지, 제41권, SP 제 6호, pp. 271-278, 2004.
- [14] M. Slaney, "A critique of pure audition," Computational Auditory Scene Analysis, 1997.

전 자 스 기



**정 명 범**

2004년 : 숭실대학교 미디어학부 졸업 (공학사)  
 2006년 : 숭실대학교 미디어학과 졸업(공학석사)  
 2006년~현재 : 숭실대학교 미디어학과 박사  
 과정  
 관심 분야 : 디지털 신호처리, 감성인식,  
 콘텐츠 공학



**성 보 경**

2006년 : 숭실대학교 미디어학부 졸업 (공학사)  
 2007년 : 숭실대학교 미디어학과 졸업(공학석사)  
 2007년~현재 : 숭실대학교 미디어학과 박사  
 과정  
 관심 분야 : 디지털 음악 분석, 음악 검색,  
 감성공학



**고 일 주**

1992년 : 숭실대학교 전산학과(공학사)  
 1994년 : 숭실대학교 전산학과(공학석사)  
 1997년 : 숭실대학교 전산학과(공학박사)  
 2003년 ~ 현재 : 숭실대학교 미디어학부 조  
 교수  
 관심 분야 : 콘텐츠, 정보검색, 감성공학