

오디오의 파형과 FFT 분석을 이용한 대표 선율 검색

(Representative Melodies Retrieval using Waveform and FFT Analysis of Audio)

정 명 범 ^{*} 고 일 주 ^{**}

(Myoung-Bum, Chung) (Il-Ju, Ko)

요 약 최근 내용 기반 음악 검색 시스템에서는 사용자의 응답 시간을 단축시키기 위해 음악의 대표 선율을 갖는 선율을 추출하여 색인하고, 검색 시 이를 사용한다. 기존 연구에서는 미디(midi) 데이터를 이용하여 대표 선율을 추출하는 방법이 제안되었으나, 미디 데이터에 한정되는 단점이 있었다. 따라서 본 논문에서는 디지털 신호처리를 이용하여 모든 오디오 파일 포맷에 적용 가능한 대표 선율 검색을 제안한다. 대표 선율 검색을 위해 FFT(Fast Fourier Transform)을 이용하여 박자와 마디를 찾고 각 마디들의 PCM 데이터로부터 높은 수치가 나타나는 빈도를 측정한다. 이때 높은 수치가 가장 많이 뭉쳐 있는 영역에서 여덟 마디 간격이 오디오 데이터의 대표 선율 영역이다. 제안 방법의 유효성을 검증하기 위한 실험으로 총 1000곡을 선택하여 대표 선율을 추출하였고, 그 결과 템포를 찾아낸 737곡 중 79.5%의 정확성을 보였다.

키워드 : 대표 선율 검색, 오디오 특징, 오디오 정보 검색, 오디오 신호 처리

Abstract Recently, we extract the representative melody of the music and index the music to reduce searching time at the content-based music retrieval system. The existing study has used MIDI data to extract a representative melody but it has a weak point that can use only MIDI data. Therefore, this paper proposes a representative melody retrieval method that can be use at all audio file format and uses digital signal processing. First, we use Fast Fourier Transform (FFT) and find the tempo and node for the representative melody retrieval. And we measure the frequency of high value that appears from PCM Data of each node. The point which the high value is gathering most is the starting point of a representative melody and an eight node from the starting point is a representative melody section of the audio data. To verify the performance of the method, we chose a thousand of the song and did the experiment to extract a representative melody from the song. In result, the accuracy of the extractive representative melody was 79.5% among the 737 songs which was found tempo.

Key words : Representative melodies retrieval, Audio Genre Classification, Audio Information Retrieval, Audio Signal Processing

* 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음

^{*} 학생회원 : 숭실대학교 미디어학과
nzin@ssu.ac.kr

^{**} 정 회 원 : 숭실대학교 미디어학과 교수
andy@ssu.ac.kr

논문접수 : 2006년 10월 16일

심사완료 : 2007년 11월 26일

: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제34권 제12호(2007.12)

Copyright©2007 한국정보과학회

1. 서 론

폭 넓은 인터넷의 확산과 정보통신기술의 발달로 사용자들은 단기간 내에 보다 풍부하고 깊이 있는 정보를 접할 수 있게 되었으며, 정보의 형식 또한 사용자들의 요구에 맞추어 이미지, 오디오, 비디오 등과 같은 멀티 미디어 데이터를 제공하게 되었다. 그러나 시시각각 늘어나는 방대한 데이터의 양으로 인해 사용자들에게는 정보 자체도 중요하지만, 이를 색인하고 검색하는 시스템이 절실히 요구되는 현실이다. 즉, 아무리 많은 데이터가 저장되어 있다 할지라도, 적시적소에 검색되어 활용될 수 없다면 가치 있는 정보라 할 수 없다.

이러한 데이터의 색인과 검색을 위해 기존에는 수작업, 텍스트 기반의 분류 시스템을 이용하였으나, 현재는 많은 분야에서 내용기반 정보 검색 시스템을 사용하고 있다. 내용기반 정보 검색 시스템은 수작업이 아닌, 정보의 내용을 수학적으로 분석하여 구조화된 기준에 따라 대표적인 특징을 추출하고, 그러한 특징을 토대로 데이터를 체계적인 구조로 색인화 및 분류를 한다. 이러한 색인화 및 분류, 검색 작업은 멀티미디어 데이터의 신속하고 정확한 정보취득을 제공할 수 있기 때문에 멀티미디어 관련 응용 연구에 필수적인 기반 기술로 사용될 수 있다.

그 중에 오디오는 영상·음향·음성 등 멀티미디어 데이터들이 공통으로 포함하고 있는 요소이다. 오디오를 분석하여 같은 오디오를 사용한 데이터나 그와 유사한 데이터를 검색 할 수 있으며, 그로부터 멀티미디어의 정서, 분위기에 따른 자동 분류 및 추천, 검색이 가능하다. 많은 연구에서 오디오 데이터의 특징 벡터들을 추출하여 장르분류, 색인 등을 하려고 노력하고 있다[1]. 그러나 오디오는 많은 양의 데이터를 가지고 있기 때문에, 전체를 분석하고 계산하려면 많은 시간이 걸린다. 기존의 오디오의 특징 벡터를 분석하는 연구에서는 오디오의 특징을 40초 이후의 구간으로부터 추출하거나, 자주 반복이 일어나는 부분을 찾아 색인하는 방법을 사용하였다[2]. 그러나 오디오 데이터의 특징 벡터는 추출 위치와 구간의 길이에 따라 다른 결과를 나타내어, 데이터마다 특징이 되는 위치를 추적하여 특징 벡터를 추출할 필요가 있다[3].

따라서 본 논문에서는 오디오 데이터의 특징 벡터 추출을 위한 대표선율 부분의 자동 추출 방법을 제안한다. 대표선율은 오디오 데이터가 나타내는 가장 특징적인 부분이며, 사람들이 가장 잘 기억하는 부분이다. 이 부분을 찾기 위해 Midi 분야에서는 오디오 데이터를 이용한 대표선율 추출 방법이 오래전부터 연구되어 왔다. 반면에 Wave와 같은 멀티미디어에 사용되는 오디오 데이터에 대한 대표선율 추출은 미비한 상태이다. 이는 Midi 분야의 대표선율 추출 방법을 오디오 데이터에 적용하기 어렵기 때문이다. 따라서 본 논문에서는 여러 가지의 대표선율 특징들 중 절정 선율을 포함한다는 것을 이용하여 오디오 데이터의 대표 선율을 추출하였다[4].

일반적으로 오디오 데이터는 도입 부분과 진행 부분, 클라이맥스 부분으로 나눌 수 있다. 이때 클라이맥스 부분이 오디오 데이터의 대표 선율이라 할 수 있으며, 그 위치는 마디별로 파형들의 높은 값이 가장 많이 모인 부분이며 절정 선율이다. 대부분의 오디오 데이터는 대표 선율이 여덟 마디의 영역으로 이루어져 있으므로, 찾아낸 위치로부터 여덟 마디의 영역이 대표 선율 부분이라 할 수 있다. 연구는 오디오 데이터의 템포(Tempo)

와 마디를 추출하고, 추출된 마디로부터 그 오디오 데이터의 가장 높은 파형의 90% 이상 되는 값들을 마디별로 측정하여 그 영역을 찾음으로부터 대표 선율을 검색할 수 있음을 제안한다.

연구에서 사용된 장르 분류 구분은 국내 오디오 데이터 서비스 업체 중 하나인 벅스 뮤직(Bugs Music)의 국내 가요 분류를 따랐으며, 찾아낸 대표 선율의 검증에 위해 논문[5]에서 정의하고 있는 반복 선율이 3번 이상 나타나는 부분이 포함된 경우 검색이 되는 것으로 결과를 측정하였다.

논문의 구성은 다음과 같다. 2장에서 내용기반 분류 및 검색 연구에 대한 관련 연구와 대표선율 추출에 관한 연구를 언급하고, 3장에서 대표선율 검색을 위한 제안 방법을 설명한다. 4장에서 제안한 대표 선율 검색에 관한 실험과 분석을 하였고, 5장에서는 결론을 제시한다.

2. 관련 연구

2.1 내용기반 분류 및 검색

기존의 내용기반 오디오 장르 분류 및 검색에 관련된 연구는 1997년 미국 Audible Magic 사의 MuscleFish [4]가 그 효시로서, 크게 디지털 신호처리(Digital Signal Processing) 기술을 이용한 방법과 미디(MIDI) 파일의 음악 표기 정보를 이용하는 2가지 방법이 있다. 디지털 신호처리를 이용한 방법은 모든 오디오 파일 포맷에 적용이 가능하며, 오디오의 음조(Pitch), 음색(Timbre), 화성(Harmony) 등의 특징을 추출하여 장르 분류, 검색을 하는 것이다. 그러나 디지털 신호처리 기법은 비교적 구현이 복잡하고, 시간이 오래 걸리는 단점이 있다. 그와 달리 미디 파일을 이용한 방법은 구현이 쉽고, 검색 시간이 짧은 장점이 있으나, 적용 범위가 미디로 한정되는 단점이 있다[6].

디지털 신호처리 기법의 논문[7]에서는 여러 가지 15초미만의 음향 효과들에 대하여 신호의 크기, 음조, 밝음, 대역폭, 하모니 등의 특징을 추출하였다. 그러한 음향 효과들의 구분으로부터 점차 연구는 오디오 데이터의 장르 분류를 하기 시작하였다. 특히 논문 [8]에서 G. Tzanetakis는 음악 장르의 자동 분류를 위해 STFT (Short Time Fourier Transform) 기반의 오디오 특성과 Wavelet Transform 기반의 리듬(Rhythm), 음조 등의 오디오 특성을 추출하여 Blues, Classical, Country, Disco, Hiphop, Jazz 등 10가지의 음악 장르 분류에서 약 70% 정도의 성공률을 보였다.

미디를 사용한 방식으로 논문[9]는 미디 파일 내에 멜로디를 이용하여 멜로디 내의 연속된 음들을 U(높음), D(낮음), S(같음)로 표현하여 데이터베이스 내에 곡들과 UDS 문자열을 비교하여 검색한다. 그러나 이 경우 문

자열 매칭을 위한 검색 속도에 대한 문제점이 있어 이를 개선하기 위한 다양한 연구가 발표되고 있다.

2.2 대표 선율 추출에 관한 연구

대표 선율은 흔히 클라이맥스, 주제 선율이라 불리며, 사용자에게 자연스럽게 인식되는 오디오의 특정 구간이다. 작곡가는 이러한 선율을 일정 범위 내에서 장식적으로 변화시키는 변주를 표현하기 때문에, 사람들은 동일한 구간으로 인식하지만, 프로그램이나 기계적으로는 서로 다른 선율로 인지하게 된다. 따라서 대표 선율을 찾기 위해 미디를 사용한 방식에서는 대표 선율이 나타내는 두 가지 특징을 이용하였다[10]. 그 첫째는 주제 선율은 일정 유사도내에서 3번 이상 반복된다는 것이다. 두 번째는 음악적 구성 요소들 중 최고 정점으로서 사용자에게 자연스럽게 인식되는 절정선율이다. 이와 같은 두 가지 특징은 실제 작곡에서도 적용되는 부분으로 문헌[5]을 참조 할 수 있다. 그 밖에 미디를 사용한 대표 선율 추출 연구로 [11,12] 등이 있다. [11]은 미디에서 정확하게 반복되는 선율정보를 주제 선율로 하여 추출하는 방법을, [12]는 음악의 내용을 최소의 의미를 갖는 단위인 동기 단위로 구분하고, 동기간의 유사성에 따라 클러스터링하여 추출하는 방법을 제안하였다.

미디를 사용한 방식은 실제적인 데이터 값이 크지 않고, 각 음표간의 간격, 음 높낮이 등이 모두 간단히 계산될 수 있기 때문에 앞선 연구들과 같이 많이 수행이 되었다. 그에 비해, 웨이브와 같이 데이터 값이 큰 오디오 데이터는 비교해야 하는 데이터의 양이 너무 많을 뿐 아니라, 기존 미디 연구에 사용한 방식을 그대로 사용하기에는 많은 무리가 있어 현재까지 많지 실험, 연구되지 않은 실정이다.

3. 대표 선율 검색을 위한 제안 방법

일반적으로 오디오 데이터는 도입 부분과 진행 부분, 클라이맥스 부분으로 나뉜다. 이때 클라이맥스 부분이 오디오 데이터의 대표 선율이며, 그 위치는 마디별로 파형들의 높은 값이 가장 많이 나타나는 부분이다. 다음의 그림 1은 제안된 대표 선율 검색을 위한 구조를 나타낸다. 대표 선율 검색을 위해서는 다음과 같이 크게 2단계로 나뉜다. 첫 번째 단계에서 오디오 데이터의 FFT (Fourier Transform) 변환 값으로부터 마디와 마디 시

작점을 찾아내며, 두 번째 단계에서 마디별 높은 파형집중도 추적을 통해 대표 선율을 찾게 된다. 이때 마디를 찾기 위한 방법은 음악의 멜로디와 리듬 중 리듬의 규칙성을 근거하여 추출한다[13]. 본 논문에서는 드럼 중에서도 가장 주기적이며 음색이 뚜렷한 스네어 드럼 소리를 기준으로 마디를 추출한다. 첫 번째 단계 오디오 데이터의 FFT 변환 값을 이용한 마디와 마디 시작점 찾기는 간략히 소개한다. 그리고 두 번째 단계 본 논문에서 제안하고자 하는 높은 파형 집중도 추적을 통한 대표 선율 검색에 대하여 설명을 한다. 파형 집중도 추적은 오디오 데이터 전체로부터 최대값을 찾은 후 그 값의 90% 이상이 되는 파형을 마디별로 측정한다. 이때 많은 집중도를 나타내는 부분이 대표 선율의 시작 영역이며, 그로부터 여덟 마디 구간이 오디오 데이터의 대표 선율이다.

3.1 오디오 데이터의 마디와 마디 시작점 찾기

마디와 마디 시작점을 찾는 방법은 논문[13]에서 제안한 방법을 이용하며, 다음은 그 방법을 간략히 설명한다.

CD 음질의 웨이브 파일은 대부분이 44.1KHz, 16Bit로 샘플링 된다. 그러한 파일의 4분 정도에 해당하는 크기는 40MByte정도이며 스테레오로 녹음되어있다. 전처리 단계에서는 이러한 데이터의 연산 회수를 줄이기 위해 스테레오 사운드를 모노로 변환한다.

대부분의 오디오 데이터들은 멜로디를 담당하는 부분과 리듬을 담당하는 부분으로 구성된다. 그리고 리듬을 구성하는 부분이 규칙성을 가진다. 여기서는 마디의 시작점과 마디를 찾기 위해 드럼파트를 구성하고 있는 스네어 드럼이라는 악기에 초점을 맞추었다. 이 악기는 대부분 곡의 리듬에서 2번째와 4번째 박자에 규칙적으로 나타나며, 표현되는 주파수가 거의 일정하다는 특성을 가진다. 일반적으로 곡들은 도입부분에서 대부분 특징을 나타내지 않음으로, 40초가 지난 다음부터 곡들의 마디 시작점을 찾는다. 일반적으로 마디의 크기는 220,000개의 샘플링 이하라는 특징을 고려하여 40초가 지난 다음부터 262,144개(1024*256)의 샘플을 채취한다.

추출한 샘플로부터 256-point FFT를 1024번 수행하며, 이 때 변환된 데이터에서 스네어 악기의 소리는 37번째 주파수 대역에서부터 43번째 주파수 대역 사이에서 잘 나타난다. FFT를 사용하여 변환된 값들로부터 스네어 악기의 소리 후보를 찾기 위해 1024의 시간을 32씩 32개의 구간으로 나누어, 37번째에서 43번째까지의 각 주파수 대역 별로 각 구간 최대값을 찾는다. 각 구간별 최대값을 정렬하여 각 주파수에서 상위 12개의 값들을 후보로 선정한 후, 선정된 값들 중 시간 위치 값이 32 간격 보다 작은 후보들을 비교하여 값이 낮은 쪽을 후보에서 삭제한다. 그림 2는 백지영의 “사랑 안 해”



그림 1 대표 선율 검색을 위한 구조도

37 - 43번째 주파수의 출현 위치													
37번째 주파		38번째 주파		39번째 주파		40번째 주파		41번째 주파		42번째 주파		43번째 주파	
765	11.699553	763	11.906539	478	11.720020	481	11.481125	771	11.538147	1023	11.635661	769	11.860615
778	11.293674	770	11.496302	0	11.672347	782	11.478787	1023	11.496082	762	11.408854	1023	11.707766
481	11.292130	478	11.391573	760	11.131529	765	11.300535	26	10.983688	787	11.299535	765	11.404283
479	11.282703	0	10.890366	26	11.053186	1023	11.284006	761	10.942732	479	11.055950	47	11.090766
27	10.860454	1023	10.706229	1023	10.848012	26	11.239510	0	10.872567	230	10.821923	479	11.093006
1023	10.791434	0	10.689262	230	10.828126	478	10.825682	477	10.787893	969	10.638306	910	10.809447
34	10.783457	936	10.658241	0	10.718169	932	10.758796	846	10.764824	945	10.656671	960	10.632283
291	10.694003	862	10.641760	856	10.658259	0	10.731441	896	10.569106	849	10.640174	22	10.618058
855	10.683558	34	10.621099	936	10.620077	856	10.678214	936	10.554173	898	10.551657	290	10.548033
936	10.631932	21	10.521324	965	10.465866	291	10.659860	910	10.519003	631	10.496951	966	10.523227
327	10.580127	325	10.391192	284	10.408996	632	10.626136	632	10.253844	26	10.473453	277	10.400086
283	10.485164	280	10.345819	910	10.396585	875	10.620636	115	10.250436	172	10.357118	326	10.331430

그림 2 37번-43번 주파수 값이 최대가 되는 12개 후보 선정

라는 곡으로부터 샘플을 추출하여 37번째에서 43번째까지의 최대값 12개 후보를 선정할 것이다.

각 주파수별로 남겨진 후보들을 다른 주파수 후보들과 비교하여, 시간 위치 차이가 ± 10 이하인 위치 값들을 찾아서 투표한 후, 위치 값들의 평균값을 구한다. 투표 결과가 6 이상인 후보들만을 추출한 후 시간 순서대로 나열하여, 후보들 사이의 시간 간격을 구한 후, 이들 중 가장 큰 값과 가장 작은 값을 뺀 나머지 값들의 평균을 구한다. 이때 구해진 평균의 2배가 우리가 찾고자 하는 마디의 길이 즉 템포이다. 그림 3의 a는 앞서 구한 12개의 후보들 중 32 간격 안에 겹치는 후보를 제외하고 남은 후보의 위치 값을 나타낸다. 그림 3의 b는 후보들로부터 얻은 값을 시간 순서대로 나열하여, 시간 간격과 시간간격 평균을 구한 것이다.

대부분의 음악에서 스네어 드럼의 주기는 2번째와 4번째 박자에 규칙적으로 나타난다. 그림 4는 KEANE의 "Everybody's Changing"라는 곡과 버즈의 "나에게로 떠나는 여행"이라는 곡에 대한 드럼을 연주하는 악보이다. 그림 4에서 선으로 테두리를 쳐진 것이 스네어 드럼

이 연주되는 부분이며 각 노래마다 규칙적으로 나는 것을 보여준다.

즉 첫 번째 스네어 드럼이 나타난 지점에서 마디 길이의 1/4 지점이 마디의 시작점이다. 따라서 마디의 시작점은 앞에서 찾은 후보들 중 가장 큰 값을 선정하고, tempo의 1/4를 뺀 값이 마디의 시작점이다. 그림 5는 FFT를 이용하여 템포와 마디 시작점, 마디 구간을 계산하는 단계를 도식화 한 것이다.

그림 5의 FFT를 이용한 템포, 마디 시작점, 마디 구간을 추출하는 것을 수식으로 나타내면 다음과 같다.

$$T = (S_2 - S_1) \times 256$$

$$(T : \text{템포}, S_1 : \text{첫 번째 스네어 드럼}, S_2 : \text{두 번째 스네어 드럼}) \tag{1}$$

$$S_p = (40 \times 44100) + \{ (S_1 \times 256) - (T/2) \}$$

(S_p : 마디 시작점) (2)

$$[N_i] = \{ S_p + 2Ti, S_p + 2T(i+1) \}$$

($[N_i]$: i 번째 마디 구간) (3)

식 (1)에서 템포를 추출하며, 식 (2)에서 템포를 이용

37 - 43번째 주파수의 출현 위치													
37번째 주파		38번째 주파		39번째 주파		40번째 주파		41번째 주파		42번째 주파		43번째 주파	
765	763	760	FLASE	771	762	769							
481	478	478	FLASE	477	479	479							
1023	1023	1023	FLASE	1023	1023	1023							
291	280	290	291	FLASE	290	290							
855	862	856	856	848	849	FLASE							
936	936	936	932	936	945	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							
FLASE	FLASE	FLASE	FLASE	FLASE	FLASE	FLASE							

후보Sort 전	765, 479, 1023, 288, 854, 936,	시간간격평균 :	122	한마디 길이 :	122*2*256 = 62464
후보Sort 후	288, 479, 765, 854, 936, 1023,	마디시작점 :	44100*40(40s) + (288-122)*256 = 1808496		
시간간격	191, 286, 89, 82, 87,				

그림 3 후보 삭제 및 시간 간격과 시간 간격 평균 계

Everybody's Changing

나에게로 떠나는 여행

그림 4 드럼 악보와 스네어 드럼이 연주되는 위치

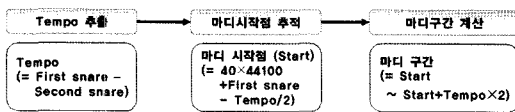


그림 5 FFT를 이용한 템포, 마디 시작점, 마디 구간 추출 순서도

하여 마디의 시작점을 찾는다. 식 (3)에서는 템포와 마디 시작점으로부터 각 마디 구간을 얻는다.

3.2 파형 집중도 추적을 통한 대표 선율 검색

본 논문에서는 3.1에서 찾은 마디와 마디 시작점, 템포를 이용하여 대표 선율을 검색한다. 오디오 데이터는 도입부분과 진행부분, 클라이맥스 부분으로 나뉜다. 이때 클라이맥스 부분이 오디오 데이터의 대표 선율이고, 대표 선율 영역은 오디오 데이터에서 전달하고자 하는 뜻을 담고 있는 부분이다. 일반적으로 그 부분은 다른 곳과 구별되고, 보다 강조가 된다. 다시 말해 오디오 데이터는 듣는 사람에게 강조하려는 부분을 강조하지 않는 부분보다 큰 소리로 표현하여 내용을 전달하는데, 그 표현 영역이 대표 선율이다. 큰 소리를 내는 것은 PCM 데이터의 수치 값이 높은 것을 의미한다. 따라서 대표 선율 영역은 높은 수치 값을 가진 PCM 데이터가 집중되어 자주 나타나는 부분이라 할 수 있다.

연구에서는 파형별 집중도를 마디로 나누어 측정하며, 집중도가 가장 높은 곳을 오디오 데이터의 대표 선율 영역 시작 지점으로 한다. 그리고 시작 지점으로부터 템포에 8을 곱한 여덟 마디가 우리가 찾아낸 대표 선율 영역이다[5].

마디별 최대 파형 집중도를 구하기 위해서는 먼저 오디오 데이터 전체로부터 파형이 가장 큰 값을 찾는다. 그 후 전 단계에서 찾은 마디의 시작 지점에서부터 각각의 마디별로 최대 파형의 90% 이상 되는 파형들의 집중도를 구한다. 이때 가장 최대 파형의 집중도가 가장 높은 마디가 찾고자 하는 대표 선율 영역의 시작 지점이다.

3.1에서 추출한 템포는 선행 연구에 의한 결과 값이다. 즉 템포가 정확히 측정이 되는 경우가 있는 반면, 그렇지 못한 경우가 있고, 비슷한 값을 측정하기도 한다. 이러한 오차율은 오디오 데이터의 후반부로 진행해 감에 따라 더욱 크게 적용된다. 따라서 오차율을 고려하여 오디오 데이터의 마디 구간이 점점 후반부로 감에 따라 각 마디별로 집중도의 가중치를 낮추어 주어야 한다. 본 논문에서는 마디 시작점으로부터 네 마디가 지날 때마다 집중도 수치를 하나씩 감소시켜 오차율을 낮춘다.

음악 계층 구조에서 악곡을 구성하는 가장 작은 단위는 동기(motif)이며, 동기 두 개가 모여서 작은악절(phrase)이 되고, 작은악절 두 개가 모여서 큰악절(period)이 된다. 음악 구성 형식에서 큰악절 두 개로 구성된 음악(두도막 형식, 8개의 동기)의 경우, 앞 큰악절의 일부분을 뒤 큰악절에서 대조를 주어 변화시키고, 뒤 큰악절에서 앞 큰악절의 일부분을 모방해 반복시켜 통일성을 갖춘다[5]. 즉, 일반적인 오디오 데이터의 대표 선율은 여덟 마디로 구성된다. 결과적으로 3.1에서 찾은 오디오 데이터의 템포와 3.2에서 얻은 대표 선율 영역 시작점으로부터 대표 선율이 나타나는 구간을 찾아낸다.

그림 6은 대표 선율 검색을 위해 만든 프로그램이며,

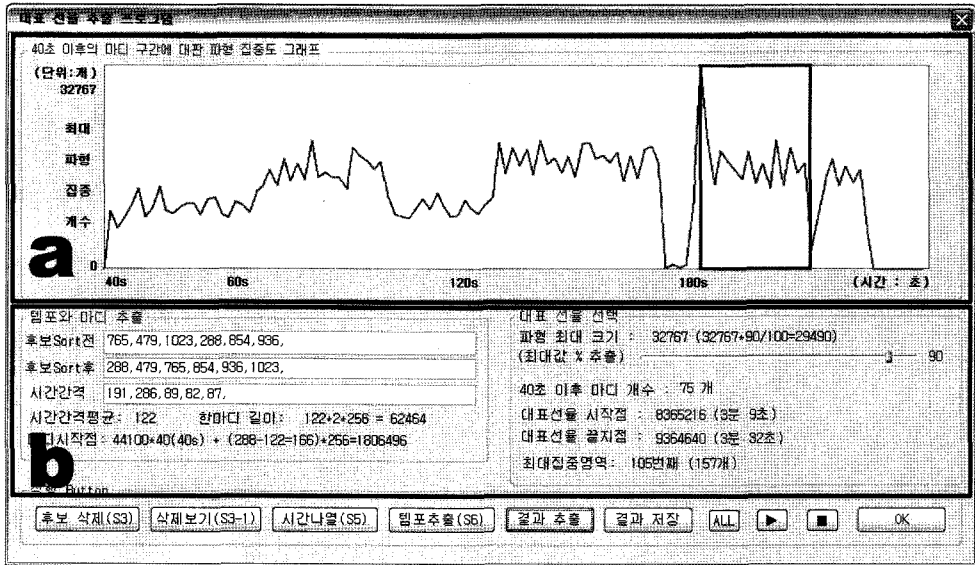


그림 6 대표 선을 영역의 검색을 위한 프로그램 샘플

a 부분은 파형이 집중해 있는 곳을 마디별 그래프로 표현하여 나타낸 것이다. a 부분 그래프에서 후반부에 그려져 있는 네모 박스가 클라이맥스 영역이며 최대 파형의 집중도가 가장 큰 곳을 시작으로 함을 알 수 있다. 그리고 그림 6의 b 부분은 앞서 언급한 템포, 마디를 추출한 값과 그로부터 실제 대표 선을 검색을 한 값을 얻은 그림이다. 실험 예로 사용된 음악은 SG 워너비의 “내 사람”이라는 곡이며, 대표 선을 검색 프로그램에 의해 얻은 결과 값은 3분 9초에서부터 3분 32초이다.

4. 대표 선을 검색 실험 및 분석

대표 선을 검색에 관한 실험은 다음과 같이 이루어졌다. Ballad, Dance, Hiphop, Rock 각각의 장르로 분류

되는 오디오 데이터 250곡씩 총 1000곡을 사용하였으며, 분류 구분은 국내 오디오 데이터 서비스 업체 중 하나인 빅스 뮤직의 국내 가요 분류를 따랐다.

또한 대표 선을 검색의 결과에 대한 정확성 여부는 논문[4,5]에서 정의하고 있는 반복 선율이 세 번 이상으로 나타나는 부분이 포함되는 경우 검색이 된 것으로 결과를 측정하였다. 표 1, 표 2, 표 3, 표 4는 각 장르별 실험 데이터와 결과 값을 각각 장르에 대하여 간략히 적은 것이다.

실험으로부터 표 5와 같은 결과를 얻을 수 있었다. 1000곡에 대해서 템포를 찾아낸 것은 737곡이었으며, 찾은 737곡 중 대표 선을 검색해 낸 것은 709곡이다. 그 중 정답을 찾아낸 곡은 586곡으로 템포가 추출된 곡

표 1 Ballad 대표 선을 검색 결과표

No	곡 명	템포	대표선을 시작부분	대표선을 끝부분	결과
1	혼자만의 겨울 (강수지)	-	-	-	-
2	몽중인 (박정현)	148	No Find	No Find	-
3	Sweet Dream (장나라)	162	3723424(1:24)	5050528(1:54)	○
4	눈분에 얼굴을 붉는다 (장나라)	135	9178272(3:28)	10284192(3:53)	○
5	언젠가는 (이상은)	180	No Find	No Find	-

표 2 Dance 대표 선을 검색 결과표

No	곡 명	템포	대표선을 시작부분	대표선을 끝부분	결과
1	나의 사랑 나의 신부 (UN)	103	3207328(1:12)	4051104(1:31)	○
2	Tonight (코요테)	168	5077408(1:55)	6453664(2:26)	○
3	No.1 (보아)	149	7648416(2:53)	8869024(3:21)	-
4	미녀와 야수 (DJ DOC)	160	3693216(1:23)	5003936(1:53)	○
5	0% (GOD)	246	6561952(2:28)	8577184(3:14)	○

표 3 Hiphop 대표 선을 검색 결과표

No	곡 명	템포	대표선을 시작부분	대표선을 끝부분	결과
1	3534 (B.Click)	402	6398112(2:25)	9691296(3:39)	○
2	Word Up (Side B)	116	4865952(1:50)	5816224(2:11)	○
3	1TYM (1TYM)	220	7966624(3:00)	9768864(3:41)	-
4	HOT 뜨거 (1TYM)	211	2166944(0:49)	3895456(1:28)	○
5	어머니 (1TYM)	117	3542688(1:20)	4501152(1:42)	○

표 4 Rock 대표 선을 검색 결과표

No	곡 명	템포	대표선을 시작부분	대표선을 끝부분	결과
1	오바엑션가면맨	113	4679328(1:46)	5605024(2:07)	○
2	부서진 입가에 머물다	60	3284640(1:14)	3776160(1:25)	-
3	인터넷 전쟁	169	4933024(1:51)	6317472(2:23)	○
4	단비	95	7650464(2:53)	8428704(3:11)	○
5	사랑은 구라파에서	-	-	-	-

표 5 대표 선을 검색 결과 및 분석표

	Ballad	Dance	Hiphop	Rock	Total
전체 곡수	250	250	250	250	1000
템포를 찾은 수	201	171	209	156	737
대표 선을 검색 수	192	166	207	144	709
정답인 수	160	142	186	98	586
정확도 (%)	79.6 %	83.0 %	89.0 %	62.8 %	79.5 %

중 79.5%의 정답률을 나타내었다. 장르별로 실험 결과를 보면 템포를 찾을 수 있는 장르는 Hip-hop이 가장 큰 정확도를 나타냄을 알 수 있다. Ballad와 Dance는 비교적 높은 정확도를 나타냈으며, Rock의 경우 Tempo를 찾지 못하여 많은 오차율을 나타내었다.

대표 선을 검색에 대한 정확도는 Hip-hop 장르에서 89%라는 높은 결과를 나타냈다. 반면 Rock 장르는 62.8%라는 낮은 결과를 얻었다. Hip-hop 장르의 정확성이 높은 이유는 규칙적인 리듬을 가지고 있고, 강조하는 부분을 제외한 음악 전체의 높낮이 차가 크지 않기 때문으로 예측된다. 리듬이 규칙적이기 때문에 템포를 찾는 실험에서도 Rock에 비해 결과 값이 높았다. 또한 파형 집중도에서도 Hip-hop 음악에 사용된 악기가 많지 않기 때문에 강조하는 부분이 더욱 뚜렷이 나타나 대표 선을 검색의 정확성을 높일 수 있었던 것으로 보인다. 반면에, Rock 장르는 Hip-hop 장르에 비해 다양한 악기를 사용하는데, 그 악기 소리들에 의해 템포를 찾는 실험에서 가장 낮은 결과 값을 얻었다. 이러한 결과는 템포를 기반으로 한 대표 선을 검색에도 영향을 미쳐 대표 선을 검색에 대한 결과 값도 가장 낮은 정확성을 보이는 것으로 예상된다. 게다가 대표 선을 검출은 노래 전 구간에서 높은 파형들이 집중된 구간을 찾는 것을 목적으로 하였다. 이러한 조건을 적용할 때 Rock 장르의 간주부분에서 연주되는 기타 소리는 종종 가장

크게 강조되어 그 부분을 대표 선을 영역으로 검색하는 오류가 발생하기도 하였다.

5. 결론과 향후 연구 방향

본 논문에서는 오디오의 파형과 FFT 분석을 이용한 대표 선을 검색 방법을 제안하였다. 앞서 관련 논문에서 관하여 언급하였듯이, 미디를 사용한 방식에서 대표 선을 추출하는 연구는 많이 이루어진 반면, 웨이브 데이터로부터 대표 선을 추출하는 연구는 미비한 실정이다. 따라서 기존 연구에 대한 비교 실험 및 분석이 불가능하다는 아쉬움이 있다. 그러나 제안 방법에 대한 효율성은 실험 결과를 통해 확인할 수 있다. 실험 데이터 중 템포를 찾은 곡은 74.7%이며, 이러한 템포를 찾은 곡에 대한 대표 선을 추출의 정확도는 79.5%로써 제안 방법이 효과가 있음을 알 수 있다. 하지만, Rock과 같은 특정 장르에 있어 높은 오류율을 나타냄을 볼 수 있다. 따라서 각 장르에 특성을 타지 않고 좀 더 높은 정확성을 얻을 수 있게 대표 선을 추출을 위한 특징을 찾아야 할 것이다.

또한 대표 선을 검색으로부터 얻은 결과 중 몇몇 개는 중복된 범위만을 찾을 뿐 대표선들의 정확한 시작 위치와 끝 위치를 찾지 못하는 오류를 나타내었다. 이를 해결하기 위해 오디오 데이터의 템포 추적에 있어 효과적인 방법을 찾아냄으로, 보다 나은 추적을 할 수 있을 것이다.

연구에서는 대표 선율의 검색을 위한 방법으로 PCM 데이터의 높은 수치를 갖는 파형의 집중도만을 사용하였다. 대표 선율의 검색에 있어 PCM 데이터의 파형 집중도 외에 음조, RMS(Root-Mean-Square) 등과 같은 요소를 추가함으로써 보다 나은 결과를 얻을 수 있을 것으로 예상된다. 앞으로의 연구는 대표 선율 검색을 위한 추가 요소를 찾을 것이며, 데이터 추적을 위한 효과적인 방법의 모색이 향후 연구 과제이다.

참 고 문 헌

- [1] 윤원중, 이강규, 박규식, "내용기반 오디오 장르 분류를 위한 신호 처리 연구", 전자공학회논문지, pp. 271-278, 2004.
- [2] 김성탁, 김상호, 김희린, 최지훈, 이한규, 홍진우, "음악 요약 생성에 관한 연구", 방송공학회논문지, pp. 3-14, 2006.
- [3] 윤원중, 이강규, 박규식, "Multi-feature Clustering을 이용한 강인한 내용 기반 음악 장르 분류 시스템에 관한 연구", 전자공학회논문지, pp. 115-120, 2005.
- [4] 구경이, 임상혁, 이재현, 김유성, "내용 기반 음악 정보 검색을 위한 음악 구성 형식을 고려한 대표 선율의 추출 및 색인", 정보처리학회논문지, 제11-D권, 제3호, 2004.
- [5] 김정양, 이종록, 김유성, "누구나 작곡할 수 있는 작곡법", 엘랜출판사, 2000.
- [6] 박상준, "기계 학습을 이용한 내용 기반의 음악 장르 분류", 공학석사학위논문, 서울대학교, 2002.
- [7] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," IEEE Multimedia, 3(2), 1996.
- [8] George Tzanetakis and Perry Cook, "Musical genre classification of audio signals," IEEE Transaction on Speech and Audio Processing 10, no.5, 293-302, 2002.
- [9] A. Ghias, J. Logan, D. Chamberlin, and B. Smith, "Query by Humming: Musical Information Retrieval in and Audio Database," ACM Multimedia, pp. 213-236, 1995.
- [10] 구경이, 신창환, 김유성, "내용 기반 음악 정보 검색에서 주제 선율의 변화 패턴을 이용한 색인 및 검색 기법", 정보과학회논문지, pp. 507-520, 2003.
- [11] C.Liu, J. Hsu, and A.L.P.chen, "Efficient Theme and Non-trivial Repeating Pattern Discovering in Music Databases," The Proceedings of the 15th International Conference on Data Engineering, pp. 14-21, 1999.
- [12] Yong-Kyoon Kang, Kyong-I Ku, and Yoo-Sung Kim, "Extracting Theme Melodies by Using a Graphical Clustering Algorithm for Content-based Music Information Retrieval," Lecture Notes in Computer Science, Springer-Verlag, pp. 84-97, 2001.
- [13] 이재원, 조찬운, 김상균, "시간 지연 신경망을 이용한 음악 장르 분류", 멀티미디어학회논문지, 제4권, 제5호,

pp. 414-422, 2001.



정 명 범

2004년 숭실대학교 미디어학부 졸업(공학사). 2006년 숭실대학교 미디어학과 졸업(공학석사). 2006년~현재 숭실대학교 미디어학과 박사과정. 관심분야는 디지털 신호처리, 감성인식, 콘텐츠 공학



고 일 주

1992년 숭실대학교 전산학과(공학사). 1994년 숭실대학교 전산학과(공학석사). 1997년 숭실대학교 전산학과(공학박사). 2003년~현재 숭실대학교 미디어학부 조교수 관심분야는 콘텐츠, 정보검색, 감성공학