

오디오 멜로디 추출 기반 특징 분석을 이용한 음악검색 방법에 관한 연구

송재종, 이석필

전자부품연구원 디지털미디어 연구센터

e-mail : jcsong@keti.re.kr

A Study on Music Retrieval method based on Audio Contents Feature Analysis

Chai-Jong Song, Sek-Phil Lee

Digital Media R&D Center, KETI

요 약

본 논문은 오디오 특징 분석을 기반으로 한 음악검색 방법에 대한 기술과 연구에 대한 내용이다. 본 연구에서는 크게 3가지의 주요 알고리즘을 이용하여 다 성음에서의 오디오 특징을 추출하고 3가지의 각자 다른 방식의 매칭 알고리즘을 기반으로 한 퓨전 매칭 방식을 제안한다. 오디오 특징으로는 메인 멜로디, 음악 구조를 분석한 세그먼트이션 정보를 이용한다. 본 연구에서 사용된 음악 DB는 음악 포털 서비스에서 제공하는 장르를 기반으로 한 8가지 장르에서 다양한 범위에서 2000곡을 선곡하였다. 오디오 특징 추출을 위한 알고리즘 개발과 매칭 알고리즘 개발을 위하여 음악 DB 2000곡 중 장르의 비율을 고려하여 100곡을 선정하고, 24명으로부터 1200개의 허밍을 녹음하였다. 24명 중 3명은 대학에서 음악을 전공하고 나머지는 음악적 교육을 받은 경험이 없는 사람들이다. 1200개의 허밍을 분석한 결과 전체 허밍 중 60%정도가 노래의 시작 부분을 허밍하거나 노래를 불렀고, 30%정도는 하이라이트 부분을 허밍 하였다. 나머지 10%정도는 자신이 가장 자신 있는 부분을 불렀다. 이러한 분석 결과를 기반으로 가장 중요한 부분은 노래가 시작되는 부분에서의 멜로디를 정확하게 찾아내는 것이 무엇보다 중요하다. 본 연구에서 검색결과의 평가는 MRR를 이용하여 측정하였다. MIDI DB를 사용한 경우가 다 성음에서 직접 멜로디를 추출한 경우보다 약간 성능이 우수하게 나왔으나 그 차이는 미미했다. 본 연구에서는 개발된 알고리즘을 이용하여 PC상에서 사용할 수 있는 클라이언트 프로그램과 Android app를 개발하였다.

1. 서론

디지털 콘텐츠가 급속도로 늘어나면서 이를 효율적으로 관리하고, 쉽게 찾아낼 수 있는 방법이 요구되고 있다. 특히, 음악과 관련된 시장은 디지털화의 영향으로 기존의 음반을 발매하던 방식에서 벗어나 음악포털 서비스에 각각의 곡으로 발표하는 방식이 일반화 되었다. 이러한 영향으로 사람들은 자신이 좋아하는 음악이나 혹은 좋아하는 가수의 노래를 찾기가 쉽지 않다. 지금까지는 음악을 검색할 땐 태그정보를 이용한 메타데이터를 검색하는 것이 일반적인 방식이지만 메타데이터를 생성하고 관리하는 방식은 시간과 비용이 만만찮게 들어간다. 또한, 음악에 대한 메타데이터 정보를 알지 못하는 경우는 원하는 곡을 찾아내는 것은 거의 불가능했다. 이러한 이유로 메타데이터 검색을 보완하는 방식으로 콘텐츠 기반 검색에 관한 요구가 꾸준히 증가하고 있는 것이 사실이다. 특히, 디지털 도서관에서는 내용기반 검색의 요구가 강하게 일어났다. 온라인 음악 포털 서비스나 디지털 도서관과 같은 분야 뿐만

아니라 노래방 기기에서도 오래 전부터 이러한 요구가 있었다. 하지만, 다성음에서 정확한 멜로디를 찾아내는 일은 쉽지 않았기 때문에 연구가 사회적 요구를 따라 주지 못했던 것도 사실이다. 최근 들어 구글에서 제공하고 있는 음성인식을 이용한 검색이 주목을 받고 있다. 이는 구글에서 텍스트기반의 검색에서 사용하던 방대한 데이터베이스를 음성인식 기술에 접목하면서 음성인식으로 한계를 한 단계 넘어섰다는 평가이다. 이와 비슷한 맥락에서 허밍기반 음악검색도 다른 연구 분야의 발전으로 한 단계 도약할 수 있는 환경이 되었다. iPhone을 시작으로 급속도로 보급량이 늘고 있는 스마트폰에도 허밍기반 검색 서비스를 제공하는 App이 존재한다. 하지만, 이들 서비스는 단순히 다른 사람의 먼저 한 허밍을 이용하여 특징 데이터베이스를 만들기 때문에 데이터의 집중현상을 해결할 방법이 없다. 이러한 서비스는 유명한 곡이나 자주 부르는 노래는 쉽게 찾아 줄 수 있으나 어려운 노래나 자주하지 않는 노래는 찾을 방법이 없다. 이러한 단점을 극복하기 위해서는 반드시 음악의 원곡에서 특징 데이터를 추출하

고 이를 이용하여 데이터베이스를 구축해야한다. 이렇게 되면 발매되는 음악의 양에 관계없이 어떠한 노래라도 원곡만 있으면 찾을 수 있다.

2. AFA 기반 오디오 검색 시스템

본 연구에서 개발하는 오디오 특징 분석 (AFA) 기반 오디오 검색 시스템의 개념도는 그림 1과 같다. 사용자가 정확한 정보를 모르고 노래 한 소절을 흥얼거리면 허밍으로부터 멜로디와 박자를 추출하여 이를 기반으로 DB를 검색하게 된다. 이 시스템의 전체 블록도를 살펴보면 데이터를 PCM형태로 입력받아 최적의 오디오 특징을 추출하기 위해 노이즈를 제거하고 입력된 신호가 허밍인지 폴리포닉 신호인지를 판별하게 된다. 이렇게 판별된 신호는 각 신호에 알맞은 오디오 특징 추출 알고리즘을 이용해 멜로디와 박자를 추출하게 된다. 추출된 특징들은 오디오 특징 DB 생성 엔진으로 보내지고 새롭게 생성된 특징을 오디오 특징 DB로 저장하게 된다. 동시에 추출된 특징은 매칭 엔진으로 보내지고 매칭엔진은 이를 가지고 오디오 특징 DB를 검색하게 된다. 검색엔진은 DB내의 복수개의 후보를 검색해 추천엔진에게 보낸다. 오디오 특징의 Description은 오디오의 고수준 특징을 기술하는 MP-QF 기술표준을 이용한다. 전체 시스템은 크게 3 블록으로 나누어진다. 입력 허밍으로부터 노이즈를 제거하고 특징을 추출하는 부분, 음악파일에서 특징을 추출하는 부분, 그리고 두 특징을 이용하여 원하는 음악을 찾는 매칭 및 추천 엔진이다. 각각의 블록은 다음과 같이 구성된다.

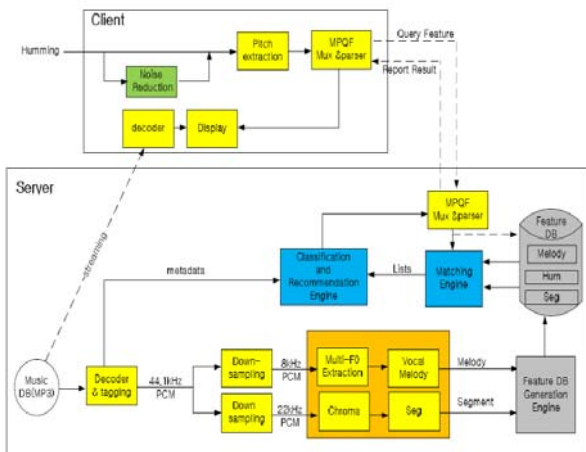


그림 1 제안하는 시스템의 전체 구성도

3. Audio Feature Extraction

음악 신호를 효율적으로 분석하고 검색하기 위해 다양한 연구가 이루어지고 있다. 입력된 다중 합성 음악 신호를 전처리 과정을 통하여 다중 피치 추출에 적합한 신호로 변

화한 후 Vocal 영역과 non-Vocal 영역으로 구분하여 처리한다. 음악에서의 vocal은 상대적으로 길기 때문에 이를 고려하여 프레임 사이즈를 음성압축기의 프레임 사이즈보다 길게 잡아야 한다. 추출된 특징들은 여러 가지 방법으로 표현 된다. 멜로디는 피치를 추적하는 멜로디 컨투어 방식을 사용한다. 정확한 노트나 스코어가 아닌 상대적인 값으로 표현함으로써 콘텐츠 기반 검색에서는 상당한 이점을 가진다. 추출되는 특징정보를 기반으로 다차원 인덱싱 DB를 생성하게 된다. 다음 그림은 다성음의 음악신호에서 주요 멜로디를 추출하는 방식을 보여주는 블록도이다. 입력 신호는 44.1kHz 스테레오를 가지는 MP3 형식의 음악파일로부터 디코딩을 통한 LPCM의 스테레오로 변환하게 된다. 이렇게 변환된 음은 음악의 특징을 추출하기 위해 필요한 4kHz 대역폭을 가질수 있도록 8kHz로 샘플링하고 이를 모노신호로 변환한다. 이렇게 변환된 신호를 이용하여 ZCR, 에너지, 주파수 영역의 피크의 편차를 이용하여 현재 프레임이 하모닉인지 아닌지를 판별하게 된다. 하모닉 프레임이라고 판별된 프레임에 한하여 다중 피치를 검사한다. 다중 피치는 주파수 영역의 피크를 이용하여 F0를 검출하고 이를 이용한 하모닉 구조를 판별한다. 하모닉 구조를 이용한 다중피치가 검출되면 이전 프레임의 피치와의 연관성을 조사하여 해당 프레임의 피치를 추출하게 된다. 추출된 피치는 연속성을 조사하여 주요 멜로디로 사용해도 되는 지를 검사하게 된다.

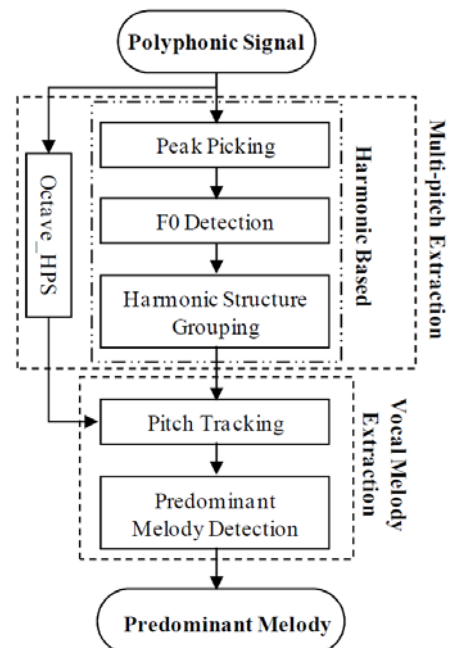


그림 2 다성음의 특징 추출 구성도

4. 허밍 특징과 음원 특징간의 패턴 매칭 엔진

오디오 특징 기반 음원 검색 알고리즘은 음원으로 부터 오디오 특징을 추출하여 오디오 특징 DB를 구축하고 입력된 음

신호를 이용하여 특징을 추출한 기존의 오디오 특징 DB와 유사성을 판별하게 된다. 일반적으로 입력 음악 정보 와 저장된 음원 사이에는 동일 음악이라고 할지라도 길이가차 발생하므로 Dynamic Time Warping를 이용한 동적 정합 매칭방법을 사용한다. 동적 정합 매칭은 서로 길이가 다른 두 패턴 간의 유사도를 측정하기 위한 방법으로 삽입과 제거 등의 방법으로 경로별 패턴 유사도를 측정할 수 있는 방법이다. DTW는 끝점 제약 조건, 최적경로 조건 그리고 격자상의 한 노드에 도달하기 위한 경로에 제한을 두어 시간상에서 지나치게 수축되거나 팽창하는 것을 막아야 한다. 또한, 서로 다른 길이를 가지는 입력 패턴과 참조 패턴간의 전 구간에 걸친 허용 가능한 영역을 제한하고, 마지막으로 국부 경로의 비용 계산 시 모두 동일한 가중치를 주지 않고 기울기에 따라서 서로 다른 가중치를 주어 시간에 비해 비합리적으로 변하는 것을 방지해야만 한다. 이렇게 정확 매칭과 유사 매칭 기법에서의 매칭 결과를 이용하여 추천엔진은 사용자 정보를 입력 받아 이를 바탕으로 사용자에게 적합한 콘텐츠 리스트를 제공하게 된다. 본 연구에서 제안하는 매칭 엔진은 DTW 방식, Linear Scale (LS), Quantized Binary Code (QBcode)방식을 이용한 퓨전 방식을 사용하였다. 각각의 알고리즘에서 최상의 매칭 결과를 가지고 세 가지의 결과를 바탕으로 최적의 결과를 찾는 방식이다. 이때 사용된 유사도 판별법은 유클리디언 디스턴스 방식을 사용하였고 선형 차이가 아닌 로그방식을 이용하여 좀 더 유사한 패턴을 찾았다.

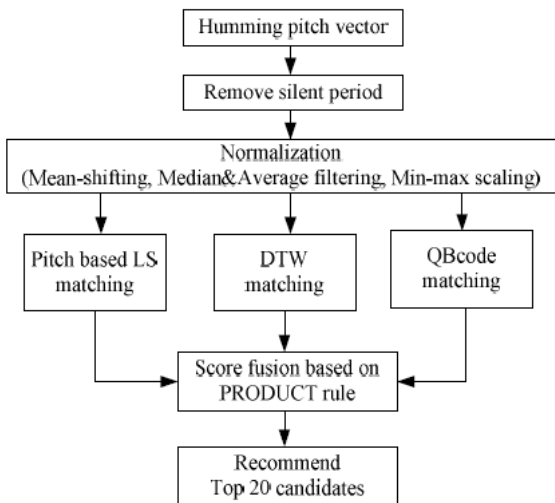


그림 4 매칭엔진 구성도

5. 결 론

본 연구에서는 오디오 특징 기반 오디오 검색 시스템을 구현하기 위한 polyphonic 음원에서 오디오 특징 추출 알고리즘, 허밍신호에서 오디오 특징 추출 알고리즘을 구현하고, 추출된 오디오 특징을 기반으로 한 다차원 오디오 특징 DB를 구축하고, 배경잡음을 제거하기 위한 알고리즘을 구현하였고, 입력된 신호에서 추출한 오디오 특징과 구축된 오디오 특징 DB사이의

다차원 인덱싱을 이용한 매칭엔진을 구현하였으며, 매칭 유사도를 이용한 사용자 추천 엔진을 구현한다. 오디오 특징 추출 알고리즘, 배경잡음 제거 알고리즘, 매칭 엔진의 구현은 상당한 수준까지 이루어 졌지만 앞으로도 더 많은 연구와 보완이 필요하다. 반면 오디오 특징 DB구현과 사용자 추천엔진은 좀더 많은 보완이 요구된다. 본 연구에서 개발된 검색 알고리즘을 검증하기 위하여 입력신호를 32ms 단위로 64ms 단위로 측정하였고 검색된 결과를 Top1, Top10, Top20를 MRR 기준으로 측정하였다. 평가에 사용된 시스템은 인텔 i7 973 코어로 8MB 메모리를 가진다. 표1은 평가결과를 나타낸다. 개발된 알고리즘과 시스템을 이용하여 PC상에서 사용가능한 어플리케이션과 Android 폰에서 사용가능한 App를 개발 하였다.

<표 1> 검색 엔진의 결과표

	Top1	Top10	Top20	MRR	Time(sec)
32ms	74.90%	89.20%	92.20%	0.793	12.4
64ms	71.10%	80.10%	83.70%	0.738	4.7



그림 5 (좌) Android App (우) PC 어플리케이션

참고문헌

[1] M. Lagrange, L. G. Martins and J. Murdoch, "Normalized cuts for predominant melodic source separation," *IEEE Trans. Audio, Speech, Langu.*, vol. 16, no. 2, Feb. 2008.

[2] 박호중, "Polyphonic 신호에서의 Vocal 영역 검출 기법 연구" 광운대학교, 2009.

[3] 김무영, "오디오 검색 시스템을 위한 배경잡음 제에에 관한 연구", 세종대학교, 2009.

[4] 박강령, "허밍 특징과 음원 특징간의 패턴 매칭 기반 음원 검색 방법 연구", 동국대학교, 2009.

[5] M. Goto, *A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station*, IEEE, Trans. Audio, Speech, and Language Processing, vol.14, no.5, Sep2006

[6] Mark A. Bartsch, *Audio Thumbnailing of Popular Music Using Chroma-Based Representations*, IEEETrans. On Multimedia, vol.7, no.1, Feb2005.