

효과적인 멜로디 표현법에 의한 고정도 오디오 콘텐츠 검색 기법

허성필*, 석수영**, 정현열***

*KT 서비스개발연구소, **일본 동북대학, ***영남대 전기·전자공학부

High Precision Audio Contents Retrieval Method by Effective Melody Representation Method

Sung-Phil Heo*, Soo-Young Suk**, Hyun-Yeol Chung***

*KT Service Development Laboratory, hsphil@kt.co.kr

** Tohoku University Graduate School of Engineering, suk@makino.ecei.tohoku.ac.jp

*** Yeungnam University School of Electrical Engineering and Computer Engineering, hychung@yu.ac.kr

요약

허밍에 의한 고정도의 오디오 정보 검색 시스템을 구현하기 위해서는 시스템 측에서 발생 가능한 문제점과 유저 측에서 발생 가능한 문제점을 함께 고려한 해결 기법이 요구된다. 유저 측에서는 허밍시 자신의 애매한 기억에 기인한 음표의 삽입이나 탈락과 같은 가상실수, 허밍 도중에 음정 및 박자의 불안정한 변화, 같은 곡을 노래 부를지라도 개인차에 의해 상이한 음정과 템포 등이 발생한다. 또한 시스템 측에서 발생 가능한 사항으로써, 비록 허밍질의가 완벽하더라도 입력 허밍 신호를 멜로디 매칭에 이용되는 정확한 특징량의 추출 및 음악 표기로의 변환이 어렵다는 점이다. 종래의 오디오 정보 검색 시스템에서는 이러한 문제점을 해결하기 위해 다양한 멜로디 표현법과 매칭 방법이 제안되고 있으나, 성능 면에서는 아직 만족할 만한 결과를 얻지 못하고 있다. 따라서 이러한 문제점들을 해결하기 위해서 본 논문에서는 허밍 멜로디의 효과적인 표현방법과 시스템 및 유저 측에서 발생 가능한 오류에 강건한 멜로디 매칭 방법을 제안한다.

1. 서론

초고속인터넷, 무선통신망의 보급과 함께 디지털기술의 발전은 음악(오디오), 영화, 애니메이션, 뮤직비디오,

게임 등과 같은 다양한 형태의 디지털콘텐츠 산업의 발전을 가져오고 있다. 지난 20세기는 네트워크 장비, 컴퓨터 등의 H/W산업이 정보통신산업의 중추를 이루었으나, 21세기에는 디지털콘텐츠에 기반을 둔 S/W산업이 세계 정보통신산업을 이끌 것으로 전망되고 있다. 따라서 소비자들의 고품질 디지털콘텐츠에 대한 수요가 확대되면서 디지털콘텐츠 산업 분야는 고부가가치를 창출하는 '황금알을 낳는 거위'로 부각되고 있다. 이미 세계 유수의 통신·방송 업체 및 H/W업체가 디지털콘텐츠 사업을 주력 사업으로 하고 있는 추세이다.

이러한 디지털 멀티미디어 콘텐츠의 사용량의 급증과 더불어, 교육, 오락, 과학 등 많은 분야에서 다양한 콘텐츠가 개발 및 제작되어 멀티미디어 데이터베이스, 디지털 TV, 인터넷 방송, 뮤직비디오 등의 응용에서 활용되고 있다. 이와 같이 방대한 양의 멀티미디어 콘텐츠를 효율적으로 사용 및 관리하기 위해서는 정보 검색(Information Retrieval) 시스템이 필요하게 된다. [1]

그러나, 이러한 멀티미디어 정보 검색 방법은 주로 텍스트와 이미지를 중심으로 발달되어 왔으나, 최근 동영상과 오디오 정보에 대한 검색 요구가 증가되고 있는 상황이다. 특히, 인터넷을 통한 온라인으로 음악울 제공하는 서비스가 늘어나고 있다. 소위 주문형 음악(MOD: Music On Demand) 혹은 주문형 오디오(AOD: Audio On

Demand)로 불리는 이들 인터넷 서비스는 방대한 분량의 음원(악곡)을 보유하고 유저의 요구에 따라 선택적으로 음악을 제공하고 있다.

일반적으로, 유저는 서비스 제공자가 특정 기준을 사용하여 미리 분류하여 정리한 웹 페이지를 선택적으로 따라 가면서 원하는 노래를 찾거나 때로는 노래 제목의 일부를 문자열 형태로 입력하여 검색함으로써 원하는 노래를 찾기도 한다. 하지만, 이러한 사전적 분류에 의한 검색 방법이나 노래 제목에 의한 검색 방법은 사용자가 찾고자 하는 노래의 제목을 미리 알고 있을 경우에만 가능하다. 또한, 검색 대상이 되는 음악 자료의 양이 방대해짐에 따라 음악 자료를 분류하기가 곤란해지고 복잡해지므로 인해, 제목 기반 검색 기법의 한계성에 부딪치게 된다.

오디오 정보 검색(Audio Information Retrieval) 방법을 살펴보면, 크게 텍스트 기반 검색(Text-based Retrieval)과 내용 기반 검색(Contents-based Retrieval)으로 분류할 수 있다. 텍스트 기반 검색은 곡명, 가수명, 가사, 작곡가, 연주가 이름 등 악곡의 특징을 나타내는 서지적인 정보를 키워드로 사용하여 검색하는 방법이다. 한편, 내용 기반에 의한 검색은 음악의 연주 내용 등을 검색키로 사용하는 것으로, 사용되는 입력정보로서 유저의 허밍을 입력키로 한 QBH(Query-by-Humming) 방법이 있다. [2], [5], [6] 본 논문은 악곡에 대한 추천율인 멜로디를 노래 부름으로써, 이것을 검색의 단서로 하여 사용자가 원하는 오디오(음원) 정보를 검색하는 방법에 대해 기술하고자 한다.

실제 환경에서 효율적으로 오디오 콘텐츠 검색 서비스를 허밍에 의해 제공하기 위해서 특히 고려되어야 할 사항으로는, 개인마다 다른 음정과 템포의 수용, 허밍시 모호한 기억에 기인하는 음표의 삽입이나 탈락과 같은 가창 실수(singing error) 등의 해결이 필요하다. 따라서, 보다 효율적이고 효과적인 멜로디 표현 방법과 이에 따른 매칭법이 요구된다.

2. 시스템 구성

오디오 정보 검색 시스템의 흐름도는 그림 1과 같이 몇 개의 주요 구성 요소 (구간 검출 모듈, 특징량 추출

모듈, 심볼 멜로디 표현 모듈, 멜로디 유사도 계산 모듈, 거리 기반 분류 모듈, 검색 결과 출력 모듈 등)로 이루어져 있다.

먼저, 유저는 자신이 검색하고자 하는 곡의 멜로디를 자유롭게 허밍한다. 시스템은 입력된 허밍으로부터 구간검출 모듈에서 음표 (note)를 검출한 후 검색에 필요한 특징량을 추출한다. 특징량 추출 모듈에서는 음장, 음고 등과 같은 음악적인 요소를 추출한다. 심볼 멜로디 표현 모듈에서는 상대화된 특징 값을 생성한다. 검색 엔진인 멜로디 유사도(similarity) 계산 모듈에서는 연속 DP (Dynamic Programming) 를 이용하여 입력 허밍과 데이터베이스 곡과의 유사성을 측정한다. 거리 기반 분류 모듈에서는 악곡 리스트를 유사도가 높은 순서로 정렬 후 곡명/유사도 값/매칭된 부분 등 악곡에 관련된 각종 정보를 출력한다.

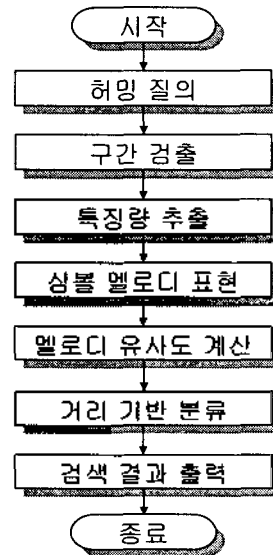


그림 1. 오디오 콘텐츠 검색 시스템의 흐름도

3. 효과적인 멜로디 표현

특징량으로써 음장은 음표 구간을 검출함으로써 얻을 수 있다. 음고의 경우는 구간검출에서 얻어진 각 음표에서 음정이 안정되어 있는 중심 프레임으로부터 피치 추출을 수행한다. 피치는 음의 높낮이를 나타내는 특성을 지니고 있다. [3] 따라서 피치가 높을수록 높은 음정을 나타낸다. 피치 추출 방법은 웨스트럼 분석(cepstrum analysis)에 의한 방법을 사용한다. [4]

실제 입력 허밍에 대해 음고 정보를 얻기 위해서는

정확한 피치 추출이 필요하다. 그러나 아무리 성능이 뛰어난 피치 추출 알고리즘이라도 옥타브의 변화가 존재하는 허밍에 대해서 실제 기본주파수가 아닌 배음을 검출하는 경우가 있다. 따라서 이러한 문제점을 해결하기 위하여 본 논문에서는 복수 피치 후보를 검색의 키로써 사용한다. [7]

입력된 허밍은 유저에 따라 템포 (tempo)와 조 (tonality)가 데이터베이스와 다르기 때문에 정규화를 수행할 필요가 있다. [2], [5] 따라서 특징량 추출 모듈로부터 동정(identification)된 음장, 음고의 계열을 전후 음표에 대해 상대음장비 (relative interval ratio), 상대음고차 (relative pitch difference)를 구해 개인차를 흡수한다. 이때, 상대음장비는 바로 이전의 음장과와의 비를 백분율(%)로 정규화하고, 상대음고차는 반음(semitone)의 차가 100센트가 되도록 표현한다.

종래의 멜로디 표현 방법은 연속된 음표간에 음장비와 음고차의 값으로 상대화를 취함으로써 특징량의 정규화를 행하였다. 그러나 이러한 멜로디 표현방법은 음표의 삽입이나 탈락이 발생하는 경우 연속적인 음표의 시계열 패턴이 변하게 된다. 따라서 이로 인해 실제 상대값의 시계열 패턴이 변화된다. 그러므로 상대값을 취하는 경우 단순히 인접하는 음표와의 상대화를 취하는 것이 아니라, 멜로디 유사도 계산에서 사용되는 연속 DP에서 매칭패스에 따라 삽입, 탈락에 대응이 가능하도록 멜로디 표현 방법을 다이내믹하게 결정할 필요가 있다. 그림 2는 데이터베이스의 어떤 곡의 1/2음표를 허밍에서 1/4음표 두개로 분리해서 허밍을 한 경우, 종래의 정적 멜로디 표현의 경우 상대값 시계열 패턴이 정답과 다르게 된다. 이러한 문제점을 해결하기 위해 전후의 음표와의 상대화뿐만 아니라, 두번째 이전의 음표와의 상대화도 고려해서 그림 3과 같이 매칭을 수행해야 한다. 이때 이러한 매칭은 DP패스의 제약 조건에 따라서 다이내믹하게 이루어 진다. [8]

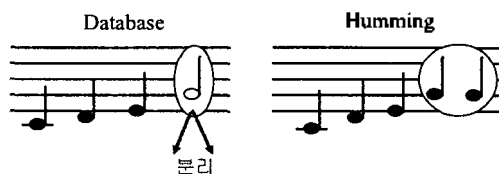


그림 2. 1/2음표가 2개의 1/4음표로 분리된 경우

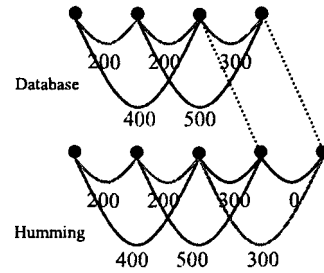


그림 3. 정적 멜로디 표현과 동적 멜로디 표현 (상대음고차 단위: cent)

4. 실험 및 평가

4.1 실험조건

실험에 사용된 오디오 데이터베이스는 주로 동요로 구성되어 있으며 전체 157곡을 사용하였다. 이것은 멜로디 라인을 가진 MIDI (Musical Instrument Digital Interface)로부터 구축하였다. 허밍 데이터는 음악에 대해 전문적인 교육을 받지 않은 5명의 사용자가 2번씩 허밍한 320곡을 사용하였다.

가창방법은 곡명만 제시하고 사용자는 마이크로폰이 있는 헤드 셋을 이용하여 자신이 알고 있는 임의의 부분 (곡의 중간 부분도 가능)의 멜로디를 자유로운 템포, 자유로운 음조로 노래를 부르게 하였다. 성능평가의 척도로서 검색 정도(retrieval accuracy)를 사용하였다.

4.2 성능평가

악곡 검색의 실험 결과는 표 1과 같다. 여기서 “Conventional”은 정적 멜로디 표현에 의한 방법을 나타내며, “Dynamic”은 제안한 다이내믹 멜로디 표현법을 사용한 결과를 나타낸다.

아래의 결과로부터, 특징량을 단순히 음장이나 음고만을 사용한 경우보다, 음장과 음고를 복합적으로 사용한 경우가 보다 좋은 검색 결과를 얻었다. 또한 다이내믹 멜로디 표현 방법이 정적 멜로디 표현 방법 보다 더 높은 검색 정도를 얻었다. 게다가 음고 특징량에 있어 복수피치후보를 사용한 경우 검색 정도는 73.9%에서 86.5%로 향상이 되었다.

따라서 제안한 다이내믹 멜로디 표현과 복수피치후보 방법을 동시에 사용한 경우 종래의 방법보다 좋은 결과를 얻은 것을 알 수가 있다.

표 1. 다양한 특징량 및 멜로디표현법에 따른 검색정도

특징량	멜로디 표현	1 위	10 위 이내
음장	Conventional	40.9 %	67.3 %
	Dynamic	42.6 %	71.2 %
음고	Conventional	61.4 %	83.9 %
	Dynamic	66.1 %	90.4 %
음장+음고	Conventional	70.8 %	86.7 %
	Dynamic	73.9 %	91.1 %
음장+음고(3) +신뢰도	Conventional	79.6 %	90.2 %
	Dynamic	86.5 %	94.1 %

5. AIR 시스템 구현

실제 GUI (Graphic User Interface)에 기초한 QBH (Query-By-Humming) 시스템인 MuseFinder를 구현하였다. 본 시스템은 MS 윈도우 플랫폼에서 Microsoft Visual C++를 이용하여 프로그래밍 하였으며 크게 5가지의 클래스로 나누어져 있다. 개략적인 클래스의 구성 요소는 음성 입/출력을 관리하는 클래스, 입력허밍 신호표시 클래스, 특징량 추출 클래스, 검색엔진 클래스, 모든 다이얼로그 관리를 위한 메인 다이얼로그 클래스로 나누어져 있다.

MuseFinder는 그림 4와 같이 3개의 서브윈도우로 구성되어 있으며, 입력허밍에 대한 각종 정보를 디스플레이해주는 사운드 정보 윈도우, 하나의 피치부터 복수의 피치 후보의 패턴과 음장정보를 함께 볼 수 있는 특징 정보 윈도우, 검색 정보 윈도우는 검색된 결과의 리스트를 보여준다. 검색 결과로써 곡명, 매칭된 지점 등을 보여주며 유저는 듣고 싶은 곡을 마우스로 클릭 함으로써 음악 감상할 수도 있다.

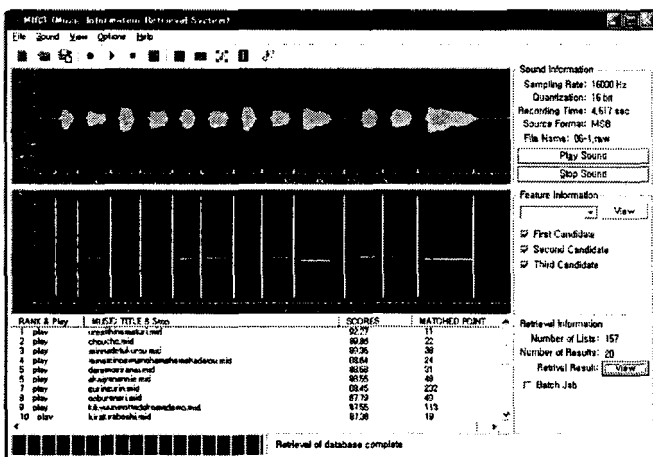


그림 4. AIR 시스템 프로토타입 (MuseFinder)

6. 결론

본 논문에서는 사용자의 입력 허밍의 오류 및 신호 처리시 발생하는 오류 등에 강한 효과적인 멜로디 표현 방법과 검색 방법을 제안하였다. 실제 사용자의 입력 허밍은 개인차에 의한 템포, 장조의 변화가 있으며, 애매한 기억에 의한 음표의 삽입/탈락 등의 여러 가지 요인으로 인해 데이터베이스 곡과 다른 입력이 들어오게 된다. 이러한 문제점을 해결하기 위해서 효과적인 멜로디 표현을 위해 최적의 상대값을 DP 패스에 따라 다이내믹하게 취하는 방법을 제안하였다. 평가실험에서 제안된 방법의 경우 top-10의 검색 정도는 94.1%의 결과를 얻었다. 이것은 종래 정적 멜로디 표현법 보다 우수한 성능을 보였다.

참고문헌

1. D. Feng, W.C.Siu and H. Zhang, "Multimedia Information Retrieval and Management: Technological Fundamentals and Applications," Springer-Verlag, 2003.
2. A. Ghias, J. Logan, D. Chamberlin and B. C. Smith, "Query By Humming: Musical Information Retrieval in an Audio Database," In Proc. ACM Multimedia, 1995.
3. L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition," Prentice-Hall, 1993.
4. Toshihiko Shirokaze, Shozo Makino and Ken'iti Kido, "Extraction of Fundamental Frequency Using Temporal Continuity over an Input Speech," Trans. IEICE Vol.73-A, No.9 pp.1537-1539, 1990.
5. Tomonari Sonoda, Makataka Goto and Yoichi Muraoka, "WWW-based Music Retrieval System," pp.343-352, In Proc. ICMC'98, 1998.
6. J. S. Roger Jang, H. Lee and J. Chen, "Super MBox: An Efficient/Effective Content-based Music Retrieval System," In the ninth ACM Multimedia Conference, pp. 636-637, 2001.
7. Sung-Phil Heo, et al., "Multiple Pitch Candidates based Music Information Retrieval Method for Query-by-Humming," AMR2003 pp. 189-200, 2003.
8. 허성필, 정현열 "허밍 질의를 이용한 오류에 강한 악곡 정보 검색 기법," 한국음향학회, 한국음향학회지, 제23권 제6호, pp. 400-409, 2004.