

YIN 피치 정보를 이용한 음악 정보 검색 시스템 구현

석수영[†], 정현열^{††}

요 약

최근 급격히 증대되고 있는 멀티미디어 데이터를 사용자에게 편하고 효과적으로 제공하는 것은 내용 기반 정보 시스템의 핵심적인 요소이다. 허밍을 이용한 음악 검색 시스템은 사용자가 찾고자 하는 음악의 선율 중 일부분을 직접 허밍으로 입력하여 데이터베이스로부터 음악을 검색할 수 있는 편리한 방법이다. 일반적인 음악 정보 검색 시스템은 고정도의 피치 검출 방법을 필요로 하고 있으나 허밍의 입력으로부터 정확한 피치 정보를 검출하기에는 어려움이 있다. 본 논문에서는 시스템의 성능 향상을 위해 기본적인 고정도 피치 정보 추출을 위해 신뢰도가 적용된 YIN 파라미터의 이용을 제안하고, 이를 적용하여 개발한 허밍을 이용한 음악 정보 검색 시스템에 대해 소개한다. 개발된 시스템은 음고, 음장 정보 및 에너지에 가중치를 두어 연속 DP 매칭을 수행하여 시스템의 성능을 향상시켰다. 성능평가를 위해 실시한 검색 실험결과 기존의 음고 정보 추출방법 중 캡스트럼 기반 다중 피치 검출 방법에 비해 본 논문에서 제안한 신뢰도를 적용한 YIN 피치 검출방법이 1위 결과에서 9.1%, 10위 결과 에서는 7.2% 성능 향상을 나타내어 제안한 방법의 유효성을 확인할 수 있었다. 또한 전체 시스템의 성능은 155곡을 대상으로 10위까지의 결과에서 92.8%의 성능을 나타내었다.

Implementation of Music Information Retrieval System using YIN Pitch Information

Soo-Young Suk[†], Hyun-Yeol Chung^{††}

ABSTRACT

Providing natural and efficient access to the fast growing multimedia information is a critical aspect for content-based information system. Query by humming system allows the user to find a song by humming part of the tune form music database. Conventional music information retrieval systems use a high precision pitch extraction method. However, it is very difficult to extract true pitch perfectly. So, In this paper, we propose to use YIN parameter with applying the reliability to reduce the pitch extraction errors. And we describes developed music information retrieval method based on a query by humming system which uses reliable feature extraction. Developed system is based on a continuous dynamic programming algorithm with features including pitch, duration and energy along with their confidence measures. The experiment showed that the proposed method could reduce the errors of the top-10 7.2% and the top-1 9.1% compared with the cepsturm based multiple pitch candidate. The overall retrieval system achieved 92.8% correct retrieval in the top-10 rank list on a database of 155 songs.

Key words: Query by Humming(허밍을 이용한 검색), Music Information Retrieval(음악 정보 검색), YIN, Pitch Detection(피치검출)

1. 서 론

음악과 인간의 감성 사이에는 아주 깊은 관계가 있다. 이는 인간은 음악을 통해 심리적 생리적 사회

적으로 많은 영향을 받고 있는 것을 의미한다. 오늘날 일상생활 속의 음악은 정보통신의 발전으로 인해 듣고 싶은 음악을 검색하고, 다른 사람에게 전송하고, 대용량으로 저장이 가능하게 되어 개인적으로 사

용하고 있는 음악 정보의 양도 급증하고 있다. 이와 같은 대용량의 음악정보를 효율적으로 처리하기 위한 연구로써 허밍을 이용한 검색(Query by humming)[1], 부분적인 음악을 이용한 검색(Query by piece of music)[2], 자동 음악 분류(Music Genre classification)[3], 음악가 식별(Artist identification)[4], 코드 검출(Chord detection)[5], 오디오 데이터의 멜로디 검출(Melody transcription from polyphonic audio)[6] 등의 다양한 연구가 시작되고 있다.

멀티미디어 정보를 위한 내용 기반 검색 기술은 일반적인 검색 정보인 가수명, 작곡가명, 제목과 같은 사전에 인지하고 있는 검색 키워드를 입력하지 않고, 음악의 일부 선율만을 입력함으로써 음악적 지식이 없는 사람도 손쉽게 사용할 수 있는 장점이 있다. 예를 들어 음악데이터베이스를 검색하기 위해 음악 자체를 녹음하여 키워드로 사용하거나 선율을 직접 입력하여 검색에 사용하고 있다. 이와 같은 내용 기반 검색 기술은 사용자에게 보다 직관적인 키워드의 입력을 제공해 줄 수 있다. 음악 검색 시스템 중에서 음악 자체를 녹음하여 키워드로 사용하는 경우에는 동일한 음악을 입력함으로써 고정도의 인식이 가능하나 키로 사용되는 음악을 녹음할 수 있는 경우에만 사용할 수 있는 단점이 있다. 따라서 언제 어디서든 원하는 음악을 검색하기 위해서는 자신이 직접 주선율을 노래로 부르는 허밍을 사용하고자 하는 연구가 진행되고 있다. 최근 허밍을 이용한 검색 시스템은 Asif Ghisa [7]에 의해 처음으로 제안이 되었으며, 이후 MELDEX [8], Themefinder [9], TuneServer [10], Super MBox [11], SoundCompass [12] 과 같은 시스템들이 개발되었다. 이런 시스템은 음악 판매점에서 신곡을 검색하거나, 노래방에서 부르고 싶은 음악을 검색할 수도 있으며, 대용량 멀티미디어 서버로부터 찾고 싶은 음악을 검색하는 등의 다양한 분야에 적용될 수 있다.

허밍을 이용한 음악 검색 시스템은 다양한 피치 추출방법과 선율 표현방법, 선율 매칭방법이 적용되고 있으며, 크게 다음과 같은 3가지의 연구부분으로 분류될 수 있다. 먼저 사용자의 허밍입력으로부터 정

확한 음고, 음장의 선율 정보 추출을 위한 연구와, 음악 데이터베이스로부터 주선율을 추출하여 정확한 정보로 저장하기 위한 연구 및 마지막으로 대용량의 데이터를 고속으로 검색하기 위한 매칭알고리즘에 대한 연구로 분류할 수 있다. 이 가운데 선율 정보를 추출하기 위해서는 사용자의 허밍입력으로부터 정확한 피치 파라미터 추출이 가장 중요하며, 시간 영역에서의 방법으로 평균 크기 차분함수(average magnitude difference function), 평균 제곱 차분 함수(average squared difference function), 유사 자기 상관 방법(similar autocorrelation methods) 등이 있으며, 주파수 영역에서의 방법으로 캡스트럼 분석을 통한 검출방법이 기본적으로 사용되고 있다. 하지만 기존의 방법은 검색시스템의 선율정보를 추출하기 위해 각 프레임마다 정확한 피치 정보를 추출하기에는 어려움이 있다. 따라서 본 논문에서는 음악검색 시스템의 성능 향상을 위해 고정도 피치 검출방법을 사용하여 음고정보 검출의 정확성을 향상시키고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 허밍을 이용한 전반적인 음악 검색 시스템에 대해 기술하고, 3장에서는 기존의 캡스트럼 기반 피치 추출방법과 제안한 YIN 기반 피치 추출방법에 대해 기술한 후 4장에서 성능평가를 위한 실험 및 그 결과에 대해 기술한 후 5장에서 결론을 맺도록 한다.

2. 허밍을 이용한 음악 검색 시스템

본 장에서는 기본적인 음악 검색 시스템에 대해 소개하고, 허밍의 입력을 노트(note) 단위의 음 구간으로 분리하는 구간 검출 방법과 추출된 파라미터를 데이터베이스 내에서 음고의 비, 음장의 비 및 음의 에너지 값을 신뢰도로 고려한 DP(Dynamic Programming) 매칭 방법에 대해 기술한다.

2.1 음악 검색 시스템의 개요

허밍을 이용한 음악 검색 시스템은 사용자가 자신이 찾고 싶은 음악 중 기억나는 부분을 직접 마이크

* 교신저자(Corresponding Author): 정현열, 주소: 경북 경산시 대동 214-1(712-749), 전화: (053)814-2496, FAX: (053)814-5713, E-mail: hychung@yu.ac.kr

접수일: 2007년 6월 7일, 완료일: 2007년 10월 5일

* 정회원, 산업기술종합연구소 정보기술연구부 음성처리

그룹(일본)

(E-mail: sy.suk@aist.go.jp)

** 정회원, 영남대학교 정보통신공학과

* 본 연구는 2006학년도 영남대학교 학술연구조성비지원으로 수행되었음

를 통해 입력함으로써 다른 방법에 비해 쉽고, 편리하게 검색이 가능하다. 그러나 개인적인 가창의 정확도, 음고저의 변화, 가창 속도의 변화에 따라 영향을 많이 받고, 입력 마이크의 특성이나 주변 환경에 따라 검색 오류가 발생할 수 있으며, 검색 대상인 음악에 비슷한 소절이 많아 입력한 가창 구간이 사용자가 실제 찾고자 하는 음악과 다른 결과로 검색되는 오류가 나타날 수 있다.

허밍을 이용한 검색 시스템의 기존 연구 예로서는 전후 피치성분과의 상대적인 변화를 UDS(Up, Down, Same)로 표현하여 검색하는 방법이 Asif Ghisa [7]에 의해 제안되었으며 국내의 경우 오동열 등[13]의 연구 예를 들 수 있다. UDS 방법은 발성의 정확성을 크게 요구하지 않으며 표현이 간편하여 고속으로 검색할 수 있는 장점이 있으나, 첫 번째 노트의 정보를 이용할 수 없고, 10개의 노트 정보로 분별할 수 있는 곡수가 2만곡도 되지 않는 단점이 있다. 또한 그림 1과 같은 유사 선율이 많이 발생하여 대용량 검색에 적합하지 않으며, 정확한 검색을 위해서는 많은 노트의 질의 정보가 필요한 단점이 있다.

UDS 표현 방법의 성능을 향상시키기 위해 Up, Down에 단계를 두어 세분화 시킨 연구가 있었으며 [14], 정확한 피치 정보를 추출할 수 있다는 가정하에 음고, 음장 정보를 이용하는 연구도 진행되고 있다 [15,16]. 이 경우 피치 정보 검출의 정확도는 크게 요구되나 음고 음장 정보를 이용함으로써 검색의 정확도를 향상시킬 수 있으며, 짧은 질의 정보로도 대용량 음악의 검색이 가능하다.

허밍을 이용한 음악 검색 시스템의 구성도는 그림 2와 같다. 구성된 시스템은 크게 전처리 및 피치 추출, 노트 구간 추출, 음악 데이터베이스로부터 메타 정보 추출 및 선형 검색 엔진의 4개의 부분으로 나눌 수 있다. 먼저 사용자의 허밍은 각 음의 "ta"로 발성하도록 하였다. 이는 무성파열음 "t"로 시작해서 유성음 "a"로 끝냄으로써 한음의 구간검출의 정도를 높일 수 있으며, 사용자 측면에서도 크게 거부감이

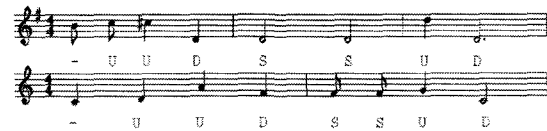


그림 1. UDS 표현을 통한 유사 선율의 예

없이 입력이 가능하다. 음성구간 검출을 통해 입력된 허밍은 전처리 과정을 통해 피치 정보를 추출한 후 하나의 음 단위로의 분리를 위해 노트 단위의 구간정보를 추출한다.

분리된 노트 정보를 기준으로 음장이 결정되며, 음고는 사전에 추출된 피치값 중 해당 노트의 중심 프레임의 값을 기준으로 추출된다. 사용된 음악 데이터베이스는 SMF(Standard Midi File)로부터 작성되었으며, 다채널 미디어정보의 경우에는 수작업으로 특정 채널을 지정함으로써 주선율정보를 추출하였다. 이때 주선율정보는 허밍의 특징과 유사하게 음장과 음고의 메타 정보로 추출하여 두며, 검색은 메타 데이터베이스로부터 입력된 음장 음고정보를 연속 DP 매칭 알고리즘을 통해 검색하여 허밍에 가장 근접한 N개의 음악을 추출한 후 사용자에게 제공하게 된다.

2.2 구간 검출

그림 3은 노트 단위로 추출된 구간 정보 추출의 예를 나타내고 있다.

그림 3의 a)는 하나의 노트가 분리된 삽입 오류를 나타내고 있으며, b)는 두 개의 노트가 결합된 삭제 오류의 예를 나타낸다. 이와 같이 노트 단위의 구간

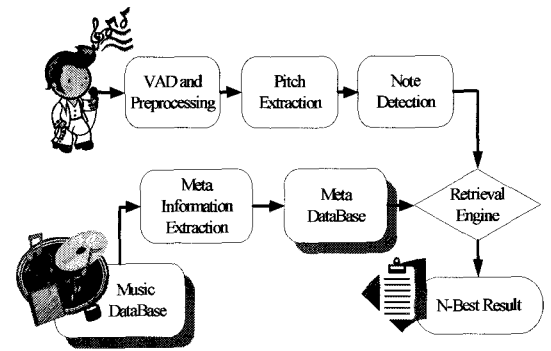


그림 2. 음악 정보 검색 시스템의 구성도

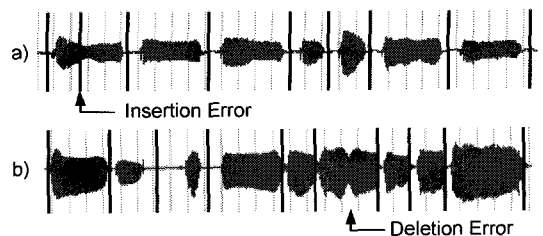


그림 3. 구간 정보 추출의 예

검출은 허밍의 정보 추출에 직접적인 영향을 주게 된다. 본 연구에서는 허밍을 600Hz~1,500Hz 대역의 밴드 패스 필터와 미분 필터를 복합적으로 사용하여 하나의 노트 단위로 구간을 분리하는 방법을 사용한다. 미분 필터의 사용은 에너지의 절대값 만으로 음을 분리시키는 경우에 비해 짧은 구간 상에서의 분해능을 향상시켜 오류를 줄일 수 있는 장점이 있다.

2.3 검색을 위한 DP 매칭

입력 허밍으로부터 추출된 음고와 음장 정보를 이용하여 메타 데이터베이스의 검색을 수행할 경우 입력된 음의 부정확함으로 인해 오류가 발생하게 된다. 이 부정확함은 구체적으로 개인적인 음조의 차이, 가창 속도의 차이, 부정확한 가창을 들 수 있으며, 이로 인해 각각 음고, 음장, 음표의 삽입과 삭제가 발생하게 된다. 이와 같은 문제점을 보완하기 위해 먼저 음고, 음장정보의 특징량을 음고, 음장의 비를 사용함으로 개인성을 제거할 수 있다. 또한 음표의 삽입과 삭제는 DP 매칭을 사용함으로 해결할 수 있다. 사용된 DP 매칭의 경로는 그림 4와 같다.

먼저 음고 정보만을 사용한 거리척도의 누적치 g_{ij} 는 식(1)와 같으며, 각 경로의 거리는 식(2)~(4)와 같다. 이때 m_i 는 i 번째 프레임의 미디 음고값을 나타내고 h_j 는 j 번째 프레임의 허밍 음고값을 나타낸다.

$$g_{ij} = \min \begin{cases} g_{i-2,j-1} + d' \\ g_{i-1,j-1} + d' \\ g_{i-1,j-2} + d'' \end{cases} \quad (1)$$

$$d = \left| \log \left\{ \frac{m_{i-1}}{m_{i-2}} \right\} - \log \left\{ \frac{h_j}{h_{j-1}} \right\} \right| / 2 \quad (2)$$

$$+ \left| \log \left\{ \frac{m_i}{m_{i-2}} \right\} - \log \left\{ \frac{h_j}{h_{j-1}} \right\} \right| / 2$$

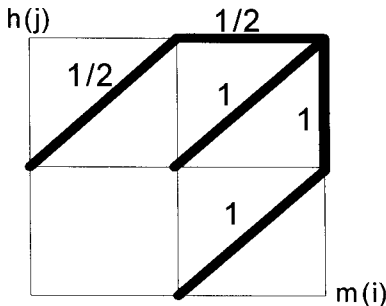


그림 4. DP 매칭의 경로

$$d' = \left| \log \left\{ \frac{m_i}{m_{i-1}} \right\} - \log \left\{ \frac{h_j}{h_{j-1}} \right\} \right| \quad (3)$$

$$d'' = \left| \log \left\{ \frac{m_i}{m_{i-1}} \right\} - \log \left\{ \frac{h_{j-1}}{h_{j-2}} \right\} \right| \quad (4)$$

$$+ \left| \log \left\{ \frac{m_i}{m_{i-1}} \right\} - \log \left\{ \frac{h_j}{h_{j-2}} \right\} \right|$$

그러나, 음고 정보만을 이용하는 경우 분리된 각 음이 동일한 정보를 제공하게 됨으로 성능의 저하를 가져온다. 이에 음구간 분리 후 각 음의 에너지값과 음장 정보의 비를 사용하여 정도를 향상 시킬 수 있다. 먼저 음의 에너지값을 사용하기 위해 거리척도 g_{ij} 는 식(5)~식(8)로 확장한다.

$$g_{ij} = \min \begin{cases} g_{i-2,j-1} + \alpha d' + (1-\alpha)p' \\ g_{i-1,j-1} + \alpha d' + (1-\alpha)p'' \\ g_{i-1,j-2} + \alpha d'' + (1-\alpha)p''' \end{cases} \quad (5)$$

$$p' = hp_i + hp_{j-1} \quad (6)$$

$$p'' = hp_j + hp_{j-1} \quad (7)$$

$$p''' = 2(hp_j + hp_{j-1} + hp_{j-2})/3 \quad (8)$$

이때 a 는 에너지값에 대한 가중치로써 a 값이 작아질수록 음고정보에 비해 에너지 정보를 많이 적용된 의미하며, hp_i 는 i 번째 프레임에서 허밍의 에너지를 나타낸다. 또한 음장정보의 비까지 포함시킨 거리척도는 식(9)~식(12)와 같다.

$$g_{ij} = \min \begin{cases} g_{i-2,j-1} + \beta(\alpha d' + (1-\alpha)p') + (1-\beta)t' \\ g_{i-1,j-1} + \beta(\alpha d' + (1-\alpha)p'') + (1-\beta)t'' \\ g_{i-1,j-2} + \beta(\alpha d'' + (1-\alpha)p''') + (1-\beta)t''' \end{cases} \quad (9)$$

$$t' = \left| \frac{mt_{i-1}}{mt_{i-2}} - \frac{ht_j}{ht_{j-1}} \right| + \left| \frac{mt_i}{mt_{i-2}} - \frac{ht_j}{ht_{j-1}} \right| \quad (10)$$

$$t'' = \left| \frac{mt_i}{mt_{i-1}} - \frac{ht_j}{ht_{j-1}} \right| \quad (11)$$

$$t''' = \left| \frac{mt_i}{mt_{i-1}} - \frac{ht_{j-1}}{ht_{j-2}} \right| + \left| \frac{mt_i}{mt_{i-1}} - \frac{ht_j}{ht_{j-2}} \right| \quad (12)$$

이때 β 값은 음장 정보에 대한 가중치로 β 값이 작아질수록 음고 및 에너지 정보에 비해 음장정보를 많이 사용함을 의미하며, a, β 값은 0~1 사이에서 실험적으로 구해진다. 전체적인 거리척도는 식(13)과 같이 나타낼 수 있다.

$$d = \beta(\alpha(\text{음고치의비}) + (1-\alpha)(\text{음의 파워값})) + (1-\beta)(\text{음장비}) \quad (13)$$

3. 피치 검출

피치의 검출은 소리의 높고 낮음을 나타내는 고유 주파수를 찾아내는 것으로 허밍의 입력으로부터 하나의 음표에 해당하는 음고를 추출하기 위해 사용된다. 추출된 음고 성분은 음장 성분과 더불어 검색의 키워드에 해당하는 것으로, 본 연구에서 이용하고 있는 음고, 음장성분의 상대치를 사용하는 시스템에서는 UDS 시스템에 비해 검색 성능에 큰 영향을 미치게 되어, 이를 정확하게 찾아내기 위한 연구가 계속 진행되어 왔다. 본 장에서는 피치검출을 위해 널리 사용되고 있는 캡스트럼 방법과 검출의 성능향상을 위해 제안하는 YIN 기반 피치추출방법에 대해 상술한다.

3.1 캡스트럼 기반 피치 검출방법

캡스트럼 기반 검출방법은 자기 상관 함수기반의 알고리즘에 비해 계산량이 작은 장점이 있으나 검출 정도가 떨어지는 단점을 가지고 있다. 이에 캡스트럼 기반 검출방법의 단점을 보완한 캡스트럼 기반 복수 피치 검출방법(MPC: Multiple Pitch Candidates)이 개발되어 좋은 성능을 나타내었다[17].

캡스트럼 기반 피치는 그림 5에서와 같은 단계를 통해 검출한다[18]. 허밍입력으로부터 전처리과정 수행 후 FFT를 이용하여 파워스펙트럼을 추출하고, 대수 IFFT를 거쳐 캡스트럼이 추출된다. 이 후 허밍의 기본주파수에 해당하는 캡스트럼의 피크값 중 최대값을 후보로 선택하고, 이를 피치 주파수로 변환한다. 이때 최대값만을 고려하는 경우 정확한 피치 주파수의 하모닉 주파수에 해당하는 피치가 선택될 가능성이 있으므로 복수 후보를 선택하여 오류를 줄일 수 있다. 즉, 복수 후보는 N 번째 최대값까지의 피치 정보를 후보로 이용하여 성능을 개선시키는 방법으로 검색 시간은 후보가 늘어날수록 선형적으로 증가하게 되는 단점이 있다. 본 연구에서는 계산량과 성능을 고려하여 3차까지의 후보를 사용하기로 한다.

그림 6은 캡스트럼 기반 MPC 검출의 예로서 a)는 입력된 허밍 파형을 나타내며, b)는 각 프레임에서 캡스트럼의 피크값중 최대값을 검출하여 주파수축에 나타낸 것이며, c)는 구간 정보 추출 후 b)의 결과로부터 해당 노트에서 중심프레임의 피치값을 노트의 음고값으로 나타낸 것이다. 또한 d)와 e)는 각각

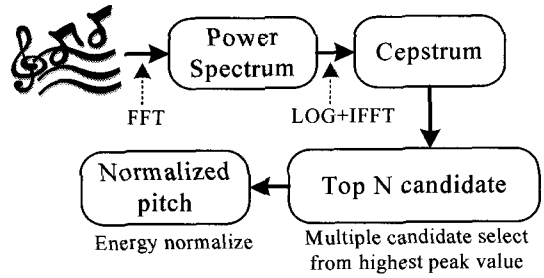


그림 5. 캡스트럼 기반 MPC 검출방법

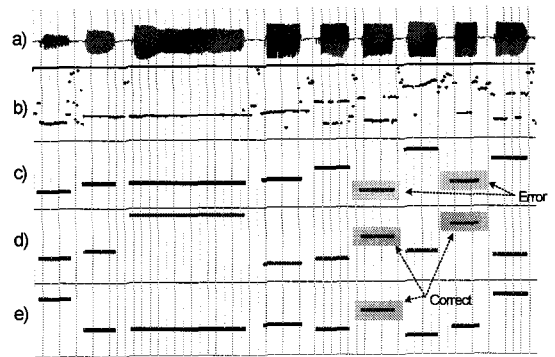


그림 6. 캡스트럼 기반 MPC 검출의 예 (a) 허밍 파형 (b) 캡스트럼 방법에 의해 추출된 프레임별 피치 정보 (c) 노트 단위로 분리된 첫 번째 음고 후보 (d) 두 번째 음고 후보 (e) 세 번째 음고 후보

해당 노트에서 중심프레임의 두 번째, 세 번째 최대값을 노트의 음고값으로 나타내었다. 그림 6 c)의 Error로 표시된 부분은 피치 추출의 오류에 따라 잘못된 음고값이 추출된 경우로써 정확한 음고값은 d)와 e)에 Correct로 표시된 음고값이다. 그림의 예에서와 같이 임의의 노트에서 피치 주파수의 고조파 성분이 검출되는 오류가 나타날 수 있으나 정확한 피치성분이 나타난 두 번째와 세 번째 후보를 사용함에 따라 성능을 향상시킬 수 있다.

3.2 YIN 기반 피치 검출 방법

일반적으로 자기상관 기반의 파라미터 추출 방법은 캡스트럼 방법에 비해 계산량이 많고, 고조파 성분이 검출될 가능성이 높다. 그러나 이런 여러 가지 오류 성분을 줄여 최적화된 YIN 방법이 de Cheveigne에 의해 제안되었다[19]. YIN 방법은 기본적인 자기상관 피치 추출방법에 비해 부가적인 파라미터의 수가 적고, 세부적인 조정이 필요 없는 장점

이 있다. YIN 이라는 이름은 자기상관과 소거 (Cancellation) 사이에서의 조화의 의미로 알고리즘의 저자가 동양 음양철학의 “Yin”과 “Yang”의 발음 으로부터 만들어 졌다. 본 연구에서는 고정도의 피치 성분을 추출하기 위해 YIN을 적용하였으며, 이에 대해 간략히 소개한다.

이산 신호 x_t 의 자기상관함수는 식(14)와 같다.

$$\gamma_t(\tau) = \sum_{j=t+1}^{t+W} x_j x_{j+\tau} \quad (14)$$

여기서 $\gamma_t(\tau)$ 는 시간 t , 지연 성분 τ 에서의 자기상관 함수이며, W 는 윈도우 크기를 나타낸다. 먼저 YIN방법은 바이어스 값에 영향을 받는 식(14)의 자기상관 함수대신에 차분함수를 이용한다.

$$d_t(\tau) = \sum_{j=t-\tau/2}^{t-\tau/2+W/2} (x_j - x_{j+\tau})^2 \quad (15)$$

식(15)의 $d_t(\tau)$ 차분함수는 0 이 되는 τ 값을 구하기 위해 τ 값이 증가할수록 윈도우폭이 감소하도록 제안 되었다. 즉 그림 7(a)에서와 같이 τ 값이 높아질수록 $d_t(\tau)$ 크기가 일정 정도 감소하게 되어 찾고자 하는 F0의 고차 성분이 검출되는 오류를 줄일 수 있는 효과가 있다. 또한 차분함수에서는 0번째가 아닌 굴곡점 중에서 최소값을 찾는 것이 목적으로 일반적으로 검색 영역을 설정해서 0번째 최소값이 검출되는 오류를 방지하고 있다. 그러나, 검색 영역은 불명확한 주기 때문에 정확한 설정이 어렵다. 이를 해결하기 위해 YIN 방법은 식(16)과 같이 누적 평균 차분함수를 사용한다.

$$d'_t(\tau) = \begin{cases} 1, & \text{if } \tau=0 \\ d_t(\tau) / \left[(1/\tau) \sum_{j=1}^{\tau} d_t(j) \right], & \text{otherwise} \end{cases} \quad (16)$$

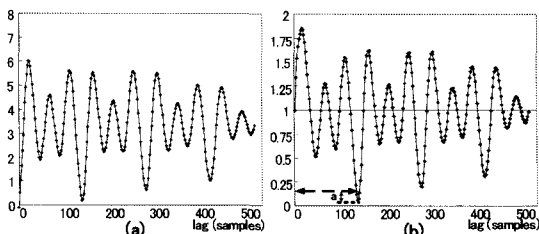


그림 7. (a) 차분함수의 예 (b) 동일한 신호에서의 누적 평균 차분함수의 예

그림 7 (b)의 예에서와 같이 누적평균 차분함수는 0번째 굴곡점이 검출되는 오류와 고조파 성분이 검출되는 오류를 줄일 수 있으며, 검색 영역을 세부적으로 지정할 필요가 없어 F0 검출영역을 넓힐 수 있는 장점이 있다.

또한 YIN 기반 피치 검출 방법의 장점으로 피치 추출과정 중 프레임별로 신뢰도 정도의 판단이 가능하며, 본 연구에는 허밍입력에 따른 노트 단위 음고 값 추출의 정확성을 향상시키기 위해 신뢰도 문턱치를 적용한 YIN 피치를 사용하였다. 신뢰도 정도의 계산은 누적평균 차분 함수의 $d_t(\tau)$ 최소값을 이용한다. 이는 허밍의 노트 단위 분리 후 나타나는 여러개의 피치 주파수중 주가 되는 음고 정보를 찾는 데 이용될 수 있다. 그림 8의 예에서 세로선으로 분리된 각 노트 단위 중 7번째 노트와 같은 경우에는 확대해서 보여주고 있는 (b) 캡스트럼 방법에 의해 추출된 피치 정보를 이용할 경우에는 하나의 노트 구간 내에서 여러개의 다른 음고값이 나타남으로 이 가운데 정확한 하나의 음고값을 결정하는데 어려움이 있다. 또한 신뢰도 문턱값을 적용하지 않은 (c) YIN 피치의 경우에도 몇 개의 음고값이 나타남을 확인할 수 있다. 그러나 신뢰도 문턱값을 적용한 (d)(e)의 경우에는 노트 구간중에 신뢰도가 낮은 YIN 피치 성분은 제거되어 (d)의 경우 2부분 (e)의 경우 하나의 음고값만이 존재하므로 노트 단위의 주 음고값 검출의 정확성을 향상시킬 수 있다. 이때 적용된 신뢰도 문턱값은 0.1에서 2.0 까지 가변적으로 적용할 수 있으며, 그림에서는 대표적인 1.0, 0.8 값을 나타내었다.

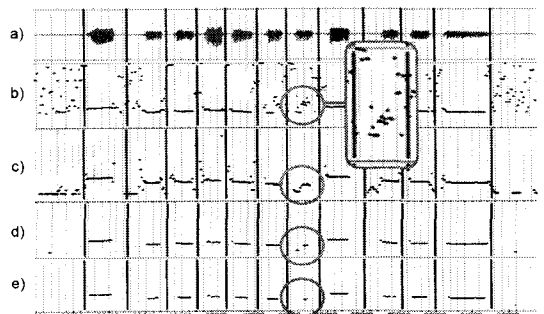


그림 8. 신뢰도가 적용된 YIN 피치 검출의 예 (a) 허밍 파형 (b) 캡스트럼방법에 의해 추출된 피치 정보 (c) YIN 방법에 의해 추출된 피치 정보 (d) 신뢰도 문턱값 1.0 이 적용된 피치 정보 (e) 신뢰도 문턱값 0.8이 적용된 피치 정보

4. 시스템 구성 및 실험 결과

4.1 시스템 구성

테스트를 위한 음악 정보 검색 시스템은 Windows Visual C++ 6.0 환경으로 개발되었다. 현재 mirs 1.4.1 버전으로 컴파일되어 있으며, 윈도우즈 실행 파일 버전은 <http://blog.naver.com/lojan/50018261528>에서 다운 받을 수 있다. 전체적인 화면 구성은 그림 9와 같다.

시스템은 전체적으로 4개의 부분으로 나뉘져 있으며, 가로블럭은 각각 위에서부터 입력 허밍의 파형, 파라미터 검출후의 음고와 음장 정보, 검색에 사용될 데이터 리스트, 검색된 결과를 나타내고 있다.

4.2 실험 결과

음악 검색 대상 데이터는 동요인 MIDI 데이터로부터 주선율을 사전에 추출한 후 검색을 위해 남성 5명으로부터 채록된 320개의 허밍 데이터를 사용하였다. 검색 데이터의 채록은 곡명을 제시한 후 자신이 알고 있는 임의의 소절을 부르도록 하였으며, 이때 실험에 사용된 허밍의 평균 노트의 수는 10.7개이며, 실험 조건은 표 1과 같다.

제한한 YIN 파라미터를 이용한 검색 성능을 비교하기 위해 캡스트럼 기반 MPC 파라미터와 비교 실험을 수행하였다. 검색의 정밀도를 향상시키기 위해 식(13)에서의 중량값 α , β 를 조정하면서 검색 성능을 비교하였다.

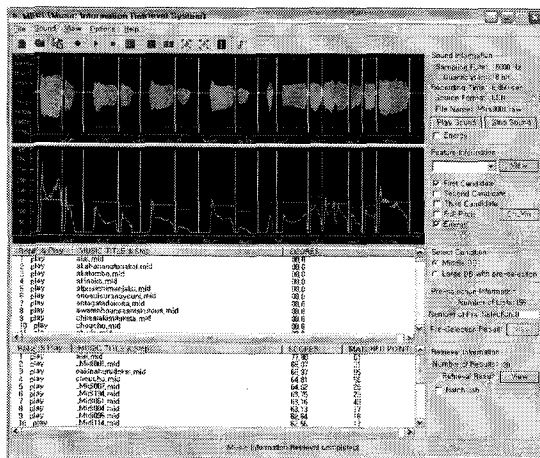


그림 9. 시스템 화면의 예

표 1. 실험 조건

악곡 데이터베이스	동요 155 곡
허밍 데이터	남성 5명의 320개
샘플링 주파수	16 kHz, 16bit
분석윈도우 분석프레임	64 ms 해밍윈도우 8ms
구간 검출	Band Pass Filter : 600-1,500Hz Differential Filter : 일차 차분필터
피치파라미터	캡스트럼, YIN

그림 10과 11에서 나타낸 바와 같이 MPC 파라미터를 이용한 검색 결과 α 값 0.8과 β 값 0.7에서 가장 좋은 성능을 나타내었으며, YIN 파라미터를 이용한 결과 α 값 0.6과 β 값 0.8에서 가장 좋은 성능을 나타내었다. 표 2는 사전실험에서 구한 최적 중량계수를 적용한 피치 정보의 검색 정도를 나타내고 있다. 캡스트럼 기반 단일 피치 후보의 경우 61.3%, MPC를 적용한 결과 68.1%에 비해 YIN 피치를 적용한 결과 75.3%를 나타내었으며, 신뢰도 문턱값 1.0을 적용한 YIN 피치의 결과 77.2%를 나타내었다.

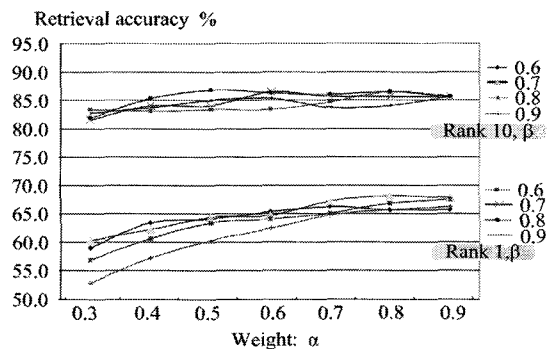


그림 10. MPC를 이용한 검색 정도

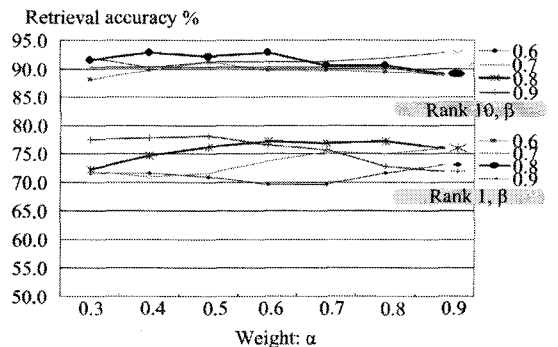


그림 11. YIN을 이용한 검색 정도

표 2. 최적의 중량 계수일 때의 검색 정도 비교

	1위	10위
캡스트럼 기반 피치	61.3%	79.38%
캡스트럼 기반 MPC	68.1%	85.6%
YIN 피치	75.3%	91.9%
신뢰도 문턱값 1.0을 적용한 YIN 피치	77.2%	92.8%

5. 결론 및 향후 연구계획

내용 기반 검색 기법 중 허밍을 이용한 음악 정보 검색 시스템은 사용자가 요구하는 음악을 허밍을 통해 마이크에 직접 입력함으로써 가수명이나 제목을 입력해야 하는 키워드 방식에 비해 전문적인 지식 없이 손쉽게 검색을 수행할 수 있는 장점이 있다. 본 논문에서는 사용자의 허밍을 분석하여 음악 정보를 검색할 수 있는 시스템을 구현한 후, 시스템의 성능을 향상시키기 위해 신뢰도 문턱값을 적용한 YIN 피치 검출 방법으로 각 노트에 해당하는 음고 정보 검출의 정확성을 향상시켰다. 또한 검색 시스템에서 사용된 DP 매칭의 거리 척도는 음고의 비 뿐만 아니라 각 음의 에너지와 음길이 정보의 비를 가중치를 두어 사용하였다. 동요 155곡을 대상으로 검색 실험을 수행한 결과 기존의 음고 정보 추출방법 중 캡스트럼 기반 다중 피치 검출 방법에 비해 제안한 신뢰도를 적용한 YIN 피치 검출방법이 1위 결과만 고려할 때 9.1%, 10위 결과까지 고려할 경우에는 7.2% 성능 향상을 나타내어 신뢰도를 적용한 YIN 피치 검출 방법의 유효성을 확인할 수 있었다.

음악 검색 시스템은 2000년 이후 활발히 연구가 진행되고 있는 분야로써, 자연스러운 허밍의 입력 및 검색 기술뿐만 아니라 MP3와 같은 음악정보로부터 주선율을 추출할 수 있는 방법이 요구되고 있으며, 대용량 음악 검색을 위해서 확률 통계적 기법을 이용한 방법[20] 및 선율 정보를 n-gram으로 표현하는 방법[21]과 같은 다양한 연구가 진행되고 있어, 향후 이를 고려한 대용량 음악 검색의 연구를 수행할 예정이다.

참고 문헌

- [1] A. Uitdenbogerd and J. Zobel, "Melodic Matching Techniques for Large Music Databases," *Proc. ACM. Multimedia*, pp. 57-66, 1999.
- [2] A.L.P. Chen, M. Chang, J. Chen, J.L. Hsu, C.H. Hsu, and S.Y.S. Hua, "Query by Music Segments: an Efficient Approach for Song Retrieval," *Proc. ICME.*, Vol.2, pp. 873-876, 2000.
- [3] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Trans. on Speech Audio Process*, Vol.10, pp. 293-302, 2002.
- [4] L. Tao and O. Mixsunori, "Music Artist Style Identification by Semi-Supervised Learning from Both Lyrics and Content," *Proc. ACM.*, pp. 364-367, 2004.
- [5] C. Namunu, "Automatic Structure Detection for Popular Music," *IEEE Trans. on Multi-Media*, Vol.13, pp. 65-77, 2006.
- [6] M. Ryyänen and A. Klapuri, "Transcription of the Singing Melody in Polyphonic Music," *Proc. ISMIR.*, pp. 222-227, 2006.
- [7] A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, "Query by Humming, Musical Information Retrieval in and Audio Database," *Proc. ACM Multimedia*, pp. 231-236, 1995.
- [8] MiDiLiB, University of Bonn, Web pages (see <http://www-mmdb.iai.uni-bonn.de/forschungsprojekte/midilib/english/>).
- [9] Themefinder, Stanford University, Web pages (see <http://www.themefinder.org/>).
- [10] TuneServer, University of Karlsruhe, Web pages (see <http://name-this-tune.com/>).
- [11] J. S. Roger Jang, H. Lee, and J. Chen, "SuperMBox: An Efficient/Effective Content-based Music Retrieval System," *Proc. ACM Multimedia*, pp. 636-637, 2001.

[12] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima, "A Practical Query-By-Humming System for a Large Music Database," *Proc. ACM Multimedia*, pp. 333-342, 2000.

[13] 오동열, 오해석, "허밍 질의 기반 음악 검색 시스템의 유사도 계산 알고리즘," *한국 컴퓨터정보학회 학회지*, 11권, 4호, pp. 137-145, 2006.

[14] Y. Zhu and M.S. Kankanhalli, "A Robust Music Retrieval Method for Query-by-Humming," *Proc. ITRE.*, pp. 89-93, 2003.

[15] 허성필, 정현열 "허밍 질의를 이용한 오류에 강한 악곡 정보 검색 기법," *한국음향학회지*, 23권, 6호, pp. 488-496. 2004.

[16] 이지연, 한민수, "허밍 운율정보를 이용한 곡목 검색 기술," *대한 음성학회 학회지*, 말소리 44. pp. 131-143, 2002.

[17] S.P. Heo, M. Suzuki, A. Ito, S. Makino, and H.Y. Chung, "Three Dimensional Continuous DP Algorithm for Multiple Pitch Candidates in Music Information Retrieval System," *Proc. ISMIR.*, pp. 235-236, 2003.

[18] S. Ahmadi and S.S. Andreas, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm," *IEEE Trans. on Speech Audio Processing*, Vol.7, pp. 333-339, 1999.

[19] A. de Cheveigne and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Jornnal ASA.*, Vol.111, 2002.

[20] J.A. Pickens, "Comparison of Language

Modeling and Probabilistic Text Information Retrieval," *Proc. ISMIR.*, pp. 111-119, 2000.

[21] J. Shifrin, B. Pardo, and W. Birmingham, "HMM-Based Musical Query Retrieval," *Proc. JCDL*. pp. 295-300, 2002.



석수영

1998년 2월 계명대학교 물리학과 (이학사)
 2000년 2월 영남대학교 멀티미디어 통신공학과(공학석사)
 2004년 2월 영남대학교 정보통신 공학과(공학박사)
 2004년 3월~2005년3월 동북대학교 전자통신공학과 연구원 박사후 과정(일본)
 2005년 4월~현재 AIST 음성정보처리 연구그룹 연구원 (일본)
 관심분야 : 음성인식, 음성처리, 멀티미디어 검색



정현열

1989년 일본 동북대학교 정보공학과(공학박사)
 1989년 3월~현재 영남대학교 전자정보공학부 교수
 1992년 7월~1993년 7월 미국 CMU Robotics 연구소 객원연구원
 2000년 6월~2000년 8월 미국 Qualcomm Inc. 수석 엔지니어
 관심분야 : 음성인식, 화자인식, 음성합성 및 DSP 응용분야