

허밍 질의를 이용한 오류에 강한 악곡 정보 검색 기법

Error-Tolerant Music Information Retrieval Method Using Query-by-Humming

허 성 필*, 정 현 열**

(Sung-Phil Heo*, Hyun-Yeol Chung**)

*KT 서비스개발연구소, **영남대학교 전기·전자공학부

(접수일자: 2004년 6월 14일; 채택일자: 2004년 8월 25일)

본고에서는 악곡정보 검색을 위한 검색키로써 허밍을 이용한 검색 방법에 관한 내용을 기술하였다. 고정도의 악곡 검색 시스템은 사용자의 불안정한 음정 및 템포 그리고 애매한 기억에 의한 음표의 삽입 및 탈락에 대응이 가능해야 한다. 또한 입력된 음향신호로부터 검색에 필요한 정확한 특징량의 추출과 효과적인 멜로디 표현 기법이 요구된다. 일반적으로 사용자의 허밍으로부터 음고 정보를 얻기 위해 입력된 음향신호로부터 피치정보를 추출하지만, 피치 추출 알고리즘은 이러한 입력 허밍으로부터 때때로 하모닉 피치를 추출한다. 이러한 문제점을 고려하여, 본 논문에서는 음고 정보의 특징량으로 복수 피치 후보를 고려한 방법을 제안한다. 게다가 복수 피치 후보에 신뢰도라는 파라미터를 도입하여 신뢰도가 높은 피치후보의 선택 가능성을 높였다. 검색엔진에서는 제안하는 복수 피치 후보의 수용을 위해 DP알고리즘을 3차원으로 확장하였다. 또한 제안하는 알고리즘은 DP패스에 따라 음표의 삽입/탈락에 다이내믹하게 대응이 가능하도록 멜로디 표현 방법을 변경하였다. 성능 평가를 위해 종래 기법과의 비교 실험 결과 보다 높은 검색 결과를 얻었다.

핵심용어: 허밍, 악곡 정보 검색, 연속 DP, 복수피치 후보, 멜로디 표현

투고분야: 음성처리 분야 (2), 음악음향 분야 (8)

This paper describes a music information retrieval system which uses humming as the key for retrieval. Humming is an easy way for the user to input a melody. However, there are several problems with humming that degrade the retrieval of information. One problem is a human factor. Sometimes people do not sing accurately, especially if they are inexperienced or unaccompanied. Another problem arises from signal processing. Therefore, a music information retrieval method should be sufficiently robust to surmount various humming errors and signal processing problems. A retrieval system has to extract pitch from the user's humming. However, pitch extraction is not perfect. It often captures half or double pitches, even if the extraction algorithms take the continuity of the pitch into account. Considering these problems, we propose a system that takes multiple pitch candidates into account. In addition to the frequencies of the pitch candidates, the confidence measures obtained from their powers are taken into consideration as well. We also propose the use of an algorithm with three dimensions that is an extension of the conventional DP algorithm, so that multiple pitch candidates can be treated. Moreover, in the proposed algorithm, DP paths are changed dynamically to take deltaPitches and IO ratios of input and reference notes into account in order to treat notes being split or unified. We carried out an evaluation experiment to compare the proposed system with a conventional system. From the experiment, the proposed method gave better retrieval performance than the conventional system.

Keywords: Humming, Music Information Retrieval, Continuous DP, Multiple Pitch Candidates, Melody Representation

ASK subject classification: Speech Signal Processing (2), Musical Acoustics and Psychoacoustics (8)

I. 서론

최근 인터넷과 무선통신의 발달과 더불어 디지털 멀티미디어 (오디오, 비디오, 음악, 이미지)의 사용량이 급증하고 있다. 이러한 멀티미디어 정보의 대두는 현대 생활에 많은 영향을 주고 있으며 특히 인간의 사고 및 미디어의 활용 방법에 커다란 변화를 주고 있다. 방대한 양의 멀티미디어 콘텐츠를 효율적으로 사용 및 관리하기 위해서는 정보 검색 시스템 (Information Retrieval System)이 필요하게 된다[1,2].

멀티미디어 정보 검색 방법은 주로 텍스트와 이미지를 중심으로 발달되어 왔으나, 최근 동영상과 오디오 정보, 특히 악곡정보에 대한 검색 요구가 증가되고 있는 상황이다. 악곡 정보 검색 (Music Information Retrieval: MIR) 방법은 크게 텍스트 기반 검색 (Text-based Retrieval)과 내용 기반 검색 (Contents-based Retrieval)으로 분류할 수가 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 사용자의 허밍을 입력키로 한 QBH (Query-by-Humming)에 대한 연구가 최근 국내외적으로 활발히 이루어지고 있는 상황이다 [3-12]. 이는 악곡에 대한 선율 정보 (멜로디)를 노래 부름으로써 이것을 검색의 단서로 하여 사용자가 원하는 악곡 정보를 검색하는 방법이다.

허밍을 이용한 MIR 시스템에 대한 본격적인 연구는 1990년대부터 이루어져 왔으며 최근 들어 이에 대한 연구가 보다 활발히 이루어지고 있는 상황이다. 예로는 Ghias 등이 제안한 QBH[4], MiDiLib[13], MELDEX[14] SuperMBox [11] 등과 같은 시스템들이 그것이다. 실제 고정도의 MIR 시스템을 구현하기 위해서 고려되어야 할

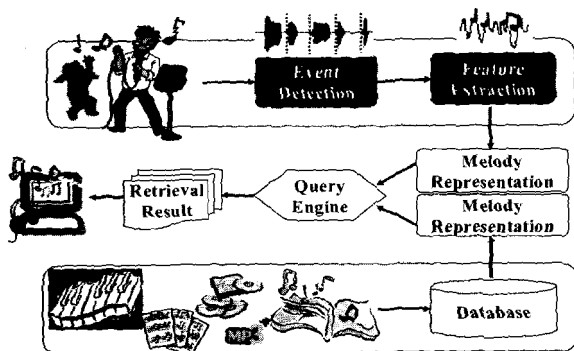


그림 1. 악곡정보 검색 시스템의 개요

Fig. 1. Overview of MIRS (music information retrieval system).

사항으로 개인마다 다른 음정과 템포의 수용, 또한 허밍 시 애매한 기억에 기인한 음표의 삽입이나 탈락과 같은 가창 실수 (singing error) 등이 있다. 또 다른 문제로서 허밍질의 (hummed query)가 완벽하더라도 입력 허밍 신호를 멜로디 매칭에 이용되는 음악 표기로 100% 정확히 변환하기가 어렵다는 점이다. 종래의 MIR 시스템에서는 다양한 멜로디 표현과 매칭 방법이 이용되고 있으나, 성능 면에서는 아직 만족할만한 결과를 얻지 못하고 있다[3,4,11]. 이러한 문제점들을 해결하기 위해서는 오류에 강한 매칭 방법과 허밍 멜로디의 효과적인 표현방법이 필요하다. 따라서 이벤트 검출, 특징 추출, 멜로디 표현, 유사도 측정 등과 같은 단계에서 기술한 문제점은 효과적이고 신뢰 할 수 있는 MIR 시스템의 구현을 위해 필수적으로 해결해야 할 과제이다.

II. MIR 시스템의 개요

제안하는 음악정보 검색 시스템은 그림 1과 같이 몇 개의 주요 구성 요소 (이벤트 검출 모듈, 특징량 추출 모듈, 멜로디 표현 모듈, 유사도 측정 모듈 등)로 이루어져 있다.

먼저, 사용자는 자신이 검색하고자 하는 곡의 멜로디를 자유롭게 허밍한다. 시스템은 입력된 허밍으로부터 이벤트 검출 모듈에서 음표 (note)를 검출한 후 검색에 필요한 특징량을 추출한다. 멜로디 표현 모듈에서는 상대화된 특징 값을 생성한다. 유사도 (similarity) 측정 모듈에서는 DP (dynamic programming) 를 이용하여 입력 허밍과 데이터베이스의 곡과 유사성을 측정한다. 검색 결과로써 악곡 리스트를 유사도가 높은 순서로 소팅 후 곡명/유사도 값/매칭 된 부분 등 악곡에 관한 각종 정보를 제시해 준다. 사용자는 검색 결과 리스트로부터 곡을 선택하여 감상을 할 수도 있다.

III. 이벤트 검출

이벤트 검출의 목적은 입력 허밍 신호로부터 한 음표의 시작점 (onset)과 끝점 (offset)의 경계를 식별하는 것이다. 진폭에 기초한 구간 검출 방법은 한 음표의 구간 검출에 적절한 임계치 (threshold value)를 설정하여

이벤트 검출을 수행한다[15]. 그러나 이러한 방법은 하나의 음표가 둘 또는 세 개의 음표로 분리되어 검출될 가능성이 있으며, 반대로 복수의 음표가 하나로 통합되어서 판단될 가능성이 있다. 따라서 이벤트 검출의 오류는 MIR 시스템의 성능에 심각한 악영향을 줄 수가 있다.

음장 정보는 이벤트 검출 과정에서 구할 수 있다. 음장은 그림 2와 같이 지속시간 (duration)과 IOI (Inter Onset Interval)으로 정의 할 수 있다. 본 논문에서, 지속시간은 어떤 음표의 시작 시간과 그 음표의 종료시간의 차를 의미한다. 한편 IOI 또는 스패 (span)은 어떤 음표의 시작시간에 대해 그 다음 음표의 시작 시간의 차를 의미한다. 예비 실험에서 음장 특징량으로써 IOI를 사용한 경우가 지속시간을 사용한 경우 보다 좋은 검색 성능을 얻었으므로 본 논문에서는 음장의 특징량으로 IOI를 사용한다[16]. 이는 실제 사용자가 노래를 부를 때 대체로 긴 음표(예를 들면 2분 음표 이상)에 대해서 실제 음장 길이만큼 정확하게 노래를 부르지는 않고 스타카토(staccato) 형태로 끊어서 노래를 부르는 이용자에 대해 보다 좋은 검색 결과를 보였다. 또한 허밍의 마지막 음표의 경우 IOI에 대한 정보가 없으므로 사용을 하지 않는 것이 좋은 결과를 보였다.

정확한 이벤트 검출을 위해서 전처리 과정으로써 대역통과필터 (Band Pass Filter: BPF) 와 차분필터 (Differential Filter: DF)를 조합한 복합 필터 처리 (Combined Filter Processing: CFP)를 그림 3과 같이 도입하였다.

그림 4는 진폭에 기초한 세그먼테이션 방법과 제안한 CFP 방법간의 이벤트 검출에 대한 비교를 보여준다.

마이크를 통해 들어온 호흡음 (respiratory sound), 섹션 A가 BPF에 의해 섹션 C와 같이 완전히 제거되었다. 또한 두개/세개의 음표가 하나의 구간으로 합쳐져서 검출되어진 영역 (섹션 B)이 차분필터의 사용으로 인

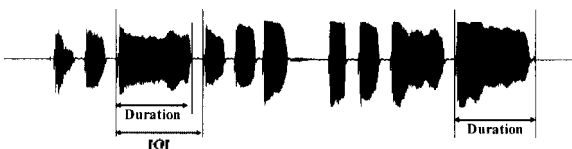


그림 2. duration 과 IOI 의 예
Fig. 2. Example of duration and IOI(Inter-Onset-Interval).

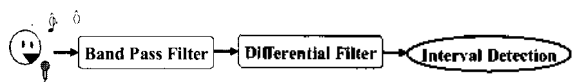


그림 3. CFP를 이용한 이벤트 검출 방법
Fig. 3. Event detection method by CFP (Combined Filter Processing).

해 섹션 D와 같이 음표가 분리되어 세그먼테이션이 이루어졌다.

IV. 특징량 추출

피치는 음의 높낮이를 나타내는 특성을 지니고 있다. 피치가 높을수록 높은 음정을 나타낸다. 따라서 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 [6,17]. 따라서 이러한 문제점을 해결하기 위하여 본 논문에서는 복수 피치 후보 (Multiple Pitch Candidates: MPC) 를 검색의 키로써 사용하는 것을 제안한다. 본 절에서는 MPC와 신뢰도 (Confidence Measure: CM) 를 얻는 방법과 피치 성능 추출 평가에 대한 내용에 대해서 기술한다.

4.1. 복수 피치 후보의 추출

피치 추출 방법은 켈스트럼 분석에 의한 방법을 사용한 대[18]. 그림 5는 피치 추출의 기본적인 흐름을 보여준다.

먼저 FFT를 이용하여 입력 신호로부터 파워 스펙트럼을 구한다. 다음으로 켈스트럼을 구하기 위해 대수와

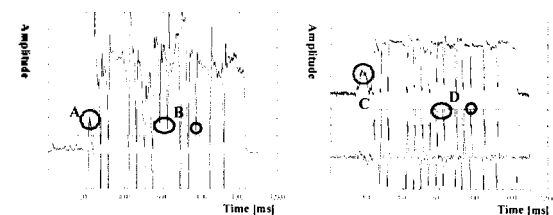


그림 4. 이벤트 검출 예 (좌: 종래의 진폭에 기초한 세그먼테이션 방법, 우: 제안한 CFP 방법)
Fig. 4. An example of the event detection (Left: conventional amplitude-based segmentation method, Right: proposed CFP method).

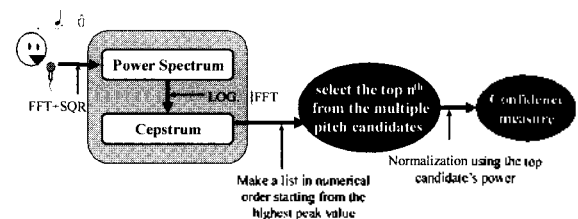


그림 5. 복수 피치 후보 및 신뢰도 추출 방법의 흐름도
Fig. 5. Extraction flow of multiple pitch candidates and confidence measure.

표 1. 피치 추출율의 평가

Table 1. Evaluation of pitches extraction rates.

순의	1위	2위 이내	3위 이내
피치 추출 정확도	88.4%	96.3%	99.7%

IFFT를 적용한다. 기본주파수 영역에 존재하는 켈스트럼 피크를 피치 후보로서 선별한다. 마지막으로, 피크의 큐브편시를 주파수 영역으로 변환한다. 특징 추출에서 얻어진 피치는 하나가 아닌 복수 피치 후보를 질의 엔진으로 보내며, 각 켈스트럼 피크의 에너지로부터 신뢰도를 계산한다. 여기서의 신뢰도는 제일 후보의 에너지 값으로 정규화한 값이다.

4.2. 피치 추출의 성능 평가

피치 추출의 정확도는 정답 피치 (correct pitch) 값과 실제 피치 추출 알고리즘에서 추출된 피치 값과의 비교에 의해 평가한다. 수작업으로 레퍼런스 데이터를 레이블링한 후 정확도 비교를 수행하였다. 피치 추출의 정확도 범위는 정답피치 (f_p)와 추출된 피치와의 관계를 그림 6과 같이 계산하였다. 본 논문에서는 추출된 피치 값이 정답 피치와 ± 20 cents 이하의 차가 날 때 정답으로 간주 하였다.

표 1은 복수의 피치 후보가 고려되었을 경우의 피치 추출의 정확도를 보여주고 있다.

오류의 원인으로써 대부분의 프레임에서 하모닉 주파수 (정답 피치에 대해 반이나 두 배의 피치값)가 추출되었다. 피치 영역이 좁으면 좁을수록 하모닉 주파수 검출에 대한 문제의 해결이 쉬워지나, 음악의 경우 피치가 한 옥타브 이상의 변화가 생기므로 하모닉 주파수의 검출에 대한 대응이 더욱 더 필요하다. 3개의 피치 후보를 이용하는 경우 피치 추출에 있어 99.7%의 정확도를 얻었다. 이러한 결과는 복수 피치 후보로써 3개의 피치 후보를 특징 벡터로서 사용하는 것이 충분하다는 것을 알 수가 있다.

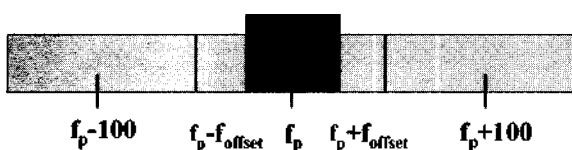


그림 6. 피치 추출 정확도의 범위
Fig. 6. Range of pitch extraction accuracy.

V. 멜로디 표현과 매칭

데이터베이스의 동일한 곡을 허밍할지라도 사용자의 허밍에 대해서 다음과 같은 점을 고려해야 한다. (1) 개인차로 인해 사용자가 허밍을 할 때 음의 높이와 빠르기가 실제 데이터베이스에 포함된 참조 패턴과 다를 수 있다. (2) 사용자의 에메모리한 기억에 의한 허밍의 경우 음표의 삽입 및 탈락 등이 발생된다. 따라서 (1)에 대해서는 허밍 데이터의 정규화가 필요하다. 일반적인 멜로디 표현 방법은 연속되는 노트에 상대적인 스펠 비 (IOratio) 와 상대적인 피치 차 (deltaPitch)를 이용하여 데이터를 정규화 한다. (2)에 대해서는 DP 매칭 알고리즘을 사용함으로써 해결을 피할 수 있다.

5.1. 정적 멜로디 표현

데이터베이스 내에 음표의 길이가 I인 어떤 악곡의 시계열 R과 입력 허밍의 음표 길이가 J인 시계열 Q에 대해 어떤 점 i, j 에 있어서의 DP 매칭 매트릭스 $g(i, j)$ 는 그림 7과 같이 구해진다. 이때 원소 Q와 R의 계산 코스트는 d로 표현한다.

만일 음표의 길이가 I 인 데이터베이스의 시계열 $R = \{i_0, i_1, i_2, \dots, i_{I-1}\}$ 과 음표의 길이가 J인 입력 허밍의 시계열 $Q = \{j_0, j_1, j_2, \dots, j_{J-1}\}$ 는 피치와 스펠의 세트 로 구성되어져 있다면, 두개의 시계열 i 와 j에 대한 유사도 $d(i, j) = |i - j|$ 는 반복적으로 계산이 가능하고 이러한 가정에 의해 유사도에 대한 거리 계산은 다음과 같은 방법에 의해서 구한다.

스텝 I, II에서의 경계조건,

$$g(0, j) = d(i_0 - j), \text{ where } 0 \leq j \leq J-1, \text{ and}$$

$$g(i, 0) = \infty, \text{ where } 0 \leq i \leq I-1$$

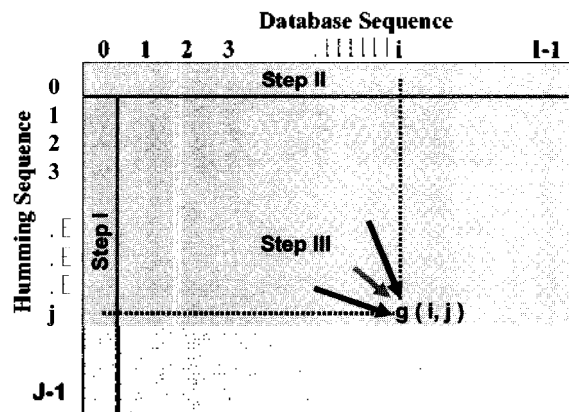


그림 7. DP 매칭 알고리즘의 예
Fig. 7. A conventional DP matching algorithm.

스텝 III에서의 일반적인 계산 절차,

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

$$d(i, j) = \gamma p(i, j) + (1 - \gamma) t(i, j). \quad (2)$$

where $i=1, \dots, I-1$ and $j=1, \dots, J-1$

이때 $d(i, j)$ 의 페널티는 그림 8과 같이 로컬 패스에 따른 가중치로 페널티를 적용한다[17].

여기서 i 와 j 는 음표의 값을 나타내고, p 는 허밍과 데이터베이스의 피치에 대한 거리 값, t 는 스펠에 대한 거리 값, γ 는 두개의 서로 다른 속성의 특징량을 통합하기 위한 가중치를 나타낸다.

p 와 t 는 각각 다음과 같이 정의한다.

$$p(i, j) = |\{m_p(i) - m_p(i-1)\} - \{h_p(j) - h_p(j-1)\}| \quad (3)$$

$$t(i, j) = |\{m_t(i) - m_t(i-1)\} - \{h_t(j) - h_t(j-1)\}| \quad (4)$$

m_p 와 m_t 는 각각 데이터베이스의 피치와 스펠 값을 나타내고, h_p 와 h_t 는 각각 입력허밍의 피치와 스펠 값을 나타낸다.

종래의 멜로디 표현 방법은 식 (2), (3)과 같이 연속된 음표간에 음장비 (IOratio)와 음고차 (deltaPitch)의 값으로 상대화를 취함으로써 특징량의 정규화를 행하였다.

상대피치 값을 센트 (cent)로 표현할 때, 센트 값은 $d(f) = 1200 (\log f - \log f_0)$ 에 의해 구할 수 있으며, 반음 (semitone)은 100센트의 값을 가진다. 여기서, f_0 는 레퍼런스 주파수를 의미한다.

그러나 종래의 멜로디 표현에는 몇 가지 문제점이 있다. 음표의 삽입이나 탈락이 발생하는 경우 연속적인 음표의 시계열 값이 변하게 되므로 이로 인해 실제 상대값의 시계열 값이 변화된다.

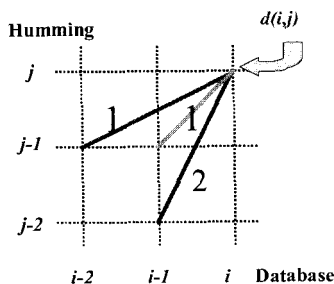


그림 8. 로컬패스와 가중치
Fig. 8. Local path constraint and weight value.

예를 들어 그림 9와 같이 음표 솔(G)이 허밍 오류에 의해 두 개의 음표로 분리 되었을 경우, 데이터베이스의 deltaPitch 열은 {200 cents, 200 cents, 300 cents}이지만 허밍으로 얻어진 deltaPitch 열은 {200 cents, 200 cents, 300 cents, 0 cents}이다. 그러므로 이 둘간의 거리는 300 cents가 된다. 따라서 이러한 비매칭은 상대적인 피치의 계산이 음표의 삽입과 탈락에 대한 가능성을 제대로 고려하지 않았기 때문이다. 그러므로 이러한 것에 대응이 가능한 멜로디 표현 방법이 필요하다.

5.2. 다이내믹 멜로디 표현

하나의 음표가 두 개로 분리가 되어 허밍이 되었을 때 현재 음표의 상대피치는 바로 전의 음표가 아닌 두번째 앞의 음표와 상대값을 취하여 매칭 할 수 있다는 가설 하에 분리된 음표의 상대피치는 (G-E=300 cents) 대신 해서 (G-G=0 cent)가 될 것이다.

이것은 상대값이 DP패스에 의해 다이내믹하게 결정될 수 있어야 된다는 것을 의미한다. 한편 탈락의 경우에도 상대 피치의 값은 위와 같은 방법으로 다이내믹하게 적용이 되어야 한다.

따라서 식(3)은 음표의 삽입과 탈락에 대응하기 위해 식(5)와 같이 변경되어야 한다.

$$\begin{aligned} p_1(i, j) &= |\{m_p(i) - m_p(i-2)\} - \{h_p(j) - h_p(j-1)\}| \\ p_2(i, j) &= |\{m_p(i) - m_p(i-1)\} - \{h_p(j) - h_p(j-1)\}| \\ p_3(i, j) &= |\{m_p(i) - m_p(i-1)\} - \{h_p(j) - h_p(j-2)\}| \end{aligned} \quad (5)$$

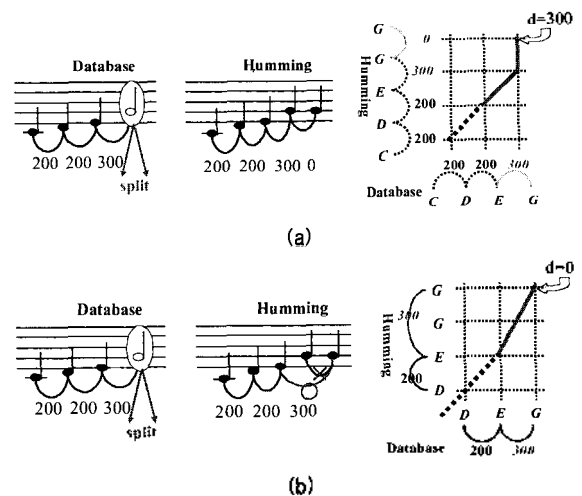


그림 9. 음표 분리가 발생한 경우 상대 피치 값의 변경 방법

(a) 종래 방법

(b) 제안 방법

Fig. 9. Relative pitch values (deltaPitch) conversion method when split occurs.

(a) Traditional method.

(b) Proposed method.

그러나 음장의 경우 단순히 상대값을 취하는 음표를 바꾸는 것만으로 문제가 해결되지 않는다. 따라서 이러한 문제점을 해결하기 위해서는 다른 접근 방법이 필요하다.

그림 10은 데이터베이스의 m3 와 m4 가 허밍시 분리되거나 합쳐지는 경우의 예를 보여준다.

데이터베이스의 m3 와 m4 가 하나의 음표로 합쳐져서 허밍된 경우 (HUM sequence A), IOratio a3/a2는 m4/m2 대신에 m3+m4/m2 와 비교되어야 한다. 마찬가지로 데이터베이스의 m3 가 두 개의 음표로 허밍이 된 경우 (HUM sequence B), b3'+b3''/b2 는 데이터베이스의 m3/m2 에 대응이 되어야 한다. 즉 음표의 삽입과 탈락에 제대로 대응하기 위해서는 음장의 시간 정보 같은 분할이나 합체가 DP패스에 의해 다이내믹하게 이루어져야 한다.

따라서 상대스팬을 정하는데 있어, 실질적인 스패는 분할과 합체의 가설에 따라 식(4)는 DP패스에 따라 다음과 같이 변경되어야 한다.

$$\begin{aligned}
 t_1(i, j) &= \left| \log \left\{ \frac{m_p(i-1) + m_p(i)}{m_p(i-2)} \right\} - \log \left\{ \frac{h_p(j)}{h_p(j-1)} \right\} \right| \\
 t_2(i, j) &= \left| \log \left\{ \frac{m_p(i)}{m_p(i-1)} \right\} - \log \left\{ \frac{h_p(j)}{h_p(j-1)} \right\} \right| \\
 t_3(i, j) &= \left| \log \left\{ \frac{m_p(i)}{m_p(i-1)} \right\} - \log \left\{ \frac{h_p(j-1) + h_p(j)}{h_p(j-2)} \right\} \right|
 \end{aligned} \tag{6}$$

최종적으로 누적 거리는 다음과 같이 변경되어야 한다.

$$g(i, j) = \min \begin{cases} g(i-2, j-1) + d_1(i, j) \\ g(i-1, j-1) + d_2(i, j) \\ g(i-1, j-2) + 2d_3(i, j) \end{cases} \tag{7}$$

$$d_e(i, j) = \gamma p_e(i, j) + (i - \gamma) t_e(i, j). \tag{8}$$

종래의 DP 매칭에서는 어떤 점 d(i, j) 에서의 값은 DP 패스에 비독립적이고 단지 i 와 j에만 의존하였다.

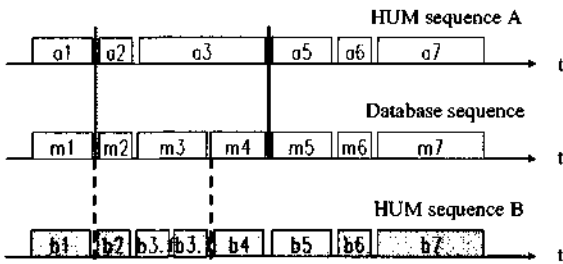


그림 10. 음표 분리 또는 통합이 발생한 경우 IOratio의 변경 방법
Fig. 10. Relative span values (IOratio) conversion method when split or union occurs.

하지만 제안한 식(8)의 경우, de=1,2,3(i, j) 는 DP 패스에 따라 계산 상대값이 변하게 된다. 이것은 유사도 측정의 계산에 있어, DP 패스에 따라 상대화를 취하는 음표값을 다이내믹하게 바꿀 수 있다는 것을 의미한다. 게다가 연속 DP는 제안하는 복수 피치 후보를 수용하기 위해 3차원으로 변경, 확장되어야 한다.

VI. 3차원 연속 DP 알고리즘

입력허밍으로부터 구한 특징량과 데이터베이스로부터 구한 특징량과는 연속 DP를 이용하여 매칭된다. 따라서 질의 엔진은 제안하는 복수 피치 후보를 수용할 수 있도록 식 (9)와 같이 3차원으로 확장할 필요성이 있다[19].

여기서 g(i, j, k) 는 입력허밍의 j번째의 음표의 k번째 피치 후보와 악곡 i번째의 음표와의 누적거리를 나타낸다.

$$g(i, j, k) = \min \begin{cases} \min \{ g(i-2, j-1, l) + d_1(i, j, k, l) \} \\ \min \{ g(i-1, j-1, l) + d_2(i, j, k, l) \} \\ \min \{ g(i-1, j-2, l) + 2d_3(i, j, k, l) \} \end{cases} \tag{9}$$

$$d_e(i, j, k, l) = \beta \{ \alpha p_e(i, j, k, l) + (1 - \alpha) C_e^{-1}(j, k, l) + (i - \beta) t_e(i, j) \}. \tag{10}$$

de=1,2,3 는 신뢰도를 가진 피치와 스패의 합산 스코어를 나타낸다. 여기서 pe(i, j, k, l)는 복수피치후보의 특정 거리, ce(j, k, l)는 신뢰도와와의 관계치이며, te(i, j) 는 스패에 대한 값을 나타낸다.

여기서 i 는 악곡의 i번째 음표 값, k 는 입력허밍의 j번째 노트의 k번째 피치후보값, l 은 입력허밍의 j번째 노트의 l번째 피치후보값을 나타낸다.

가중치 α 와 β 는 특징량 (피치, 신뢰도, 스패)의 상대적 기여도를 반영하기 위해 다양하게 변경 할 수 있는 파라미터이다.

즉, α =1 이고 β =1 일 때, 가중 기여도는 단지 피치에만 근거하고, 반면 β 가 0일 때, 가중 기여도는 스패만을 특징량으로 한 경우의 유사도를 얻을 수 있다.

pe, ce 와 te 는 각각 다음과 같이 정의한다.

$$\begin{aligned}
 p_1(i, j, k, l) &= |\{m_p(i) - m_p(i-2)\} - \{h_p(j, k) - h_p(j-1, l)\}| \\
 p_2(i, j, k, l) &= |\{m_p(i) - m_p(i-1)\} - \{h_p(j, k) - h_p(j-1, l)\}| \\
 p_3(i, j, k, l) &= |\{m_p(i) - m_p(i-1)\} - \{h_p(j, k) - h_p(j-2, l)\}|
 \end{aligned} \tag{11}$$

$$\begin{aligned}
 c_1(j, k, l) &= |(h_c(j, k) - h_c(j-1, l))| \\
 c_2(j, k, l) &= |(h_c(j, k) - h_c(j-1, l))| \\
 c_3(j, k, l) &= |(h_c(j, k) - h_c(j-2, l))|
 \end{aligned}
 \tag{12}$$

$$\begin{aligned}
 t_1(i, j) &= \left| \log \left\{ \frac{m_c(i-1) + m_c(i)}{m_c(i-2)} \right\} - \log \left\{ \frac{h_c(j)}{h_c(j-1)} \right\} \right| \\
 t_2(i, j) &= \left| \log \left\{ \frac{m_c(i)}{m_c(i-1)} \right\} - \log \left\{ \frac{h_c(j)}{h_c(j-1)} \right\} \right| \\
 t_3(i, j) &= \left| \log \left\{ \frac{m_c(i)}{m_c(i-1)} \right\} - \log \left\{ \frac{h_c(j-1) + h_c(j)}{h_c(j-2)} \right\} \right|
 \end{aligned}
 \tag{13}$$

여기서 m_c^* 과 h_c^* 는 데이터베이스와 입력 허밍의 시계열을 나타낸다. 식 (11), 식 (12) 그리고 식 (13)의 h_p , h_c , 와 h_t 는 각각 복수피치후보, 신뢰도, 스펠 값을 나타낸다.

그림 11은 복수피치후보를 가진 입력허밍이 3차원으로 확장된 DP 평면에서 실제 데이터베이스와 매칭이 이루어지는 메커니즘을 보여주고 있다. 허밍과 데이터베이스와의 매칭을 살펴보면, 제안된 알고리즘은 피치후보점의 모든 후보점과의 조합을 계산하고 최적의 후보점과 패스를 선택 결정한다.

VII. 실험 및 결과

7.1. 실험 조건

실험에 사용된 악곡DB는 주로 동요로 구성되어 있으며 전체 155곡을 사용하였다. 이것은 멜로디 라인을 가진 MIDI (Musical Instrument Digital Interface)로부터 구축하였다. 허밍 데이터는 음악에 대해 전문적인 교육을 받지 않은 5명의 이용자가 2번씩 허밍한 320곡을 사용하였다.

가창방법은 곡명만 제시하고 사용자는 마이크론이 있는 헤드 셋을 이용하여 자신이 알고 있는 임의의 부분(곡의 중간 부분도 가능)의 멜로디를 자유로운 템포, 자

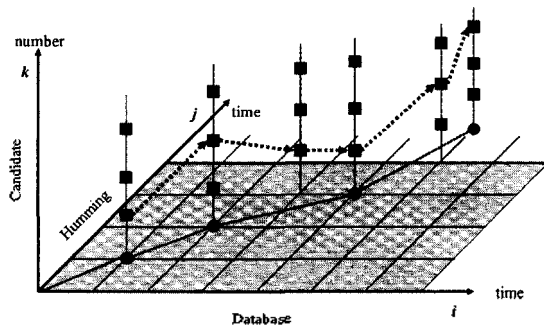


그림 11. 3차원 연속 DP알고리즘을 이용한 매칭 흐름의 예
Fig. 11. Example of the matching flow using the three-dimensional continuous DP algorithm.

표 2. 실험 조건

Table 2. Experimental conditions.

악곡 데이터베이스	동요 155곡
허밍 데이터	5명이 동일곡을 2번씩 허밍
샘플링 주파수	16 kHz
분석 윈도우 분석 프레임	64 ms 해밍 윈도우 8 ms
구간 검출	BPF: 600~1500 Hz DF: 일차 차분 필터
특징량	음장, 복수피치후보, 신뢰도

유로운 음조로 노래를 부르게 하였다. 표 2는 실험 조건을 보여 주고 있다. 허밍의 평균 음표 수는 9.8 음표였고, 평균 허밍 시간은 4.6초였다. 성능평가의 척도로서 검색 정도 (retrieval accuracy)를 사용하였다.

7.2. 실험 결과

제안한 방법의 성능을 비교 평가하기 위해 종래의 악곡 검색 시스템을 구현하여 서로의 성능을 비교하였다. 악곡 검색의 실험 결과는 표 3과 같다. 여기서 "Coarse-to-Fine" 와 "Category 27" 는 참고문헌 [5]의 방법을 나타낸다. 또한 "Conventional"은 5.1절의 정적 멜로디 표현에 의한 방법을 나타내며 "Dynamic"은 제안한 다이내믹 멜로디 표현법을 사용한 결과를 나타낸다.

위의 결과로부터, 특징량을 단순히 스펠이나 피치만을 사용한 경우보다, 스펠과 피치를 복합적으로 사용한 경우가 보다 좋은 검색 결과를 얻었다. 또한 다이내믹 멜로디 표현 방법이 정적 멜로디 표현 방법 보다 3-7포인트 더 높은 검색 정도를 얻었다. 게다가 복수피치후보를 사용함으로써 검색정도는 73.9%에서 86.5%로 향상이 되었다.

또한 종래 MIR 시스템의 "Coarse-to-Fine" 과 "Category 27" 의 경우 1위에 랭크된 곡의 결과가 각각

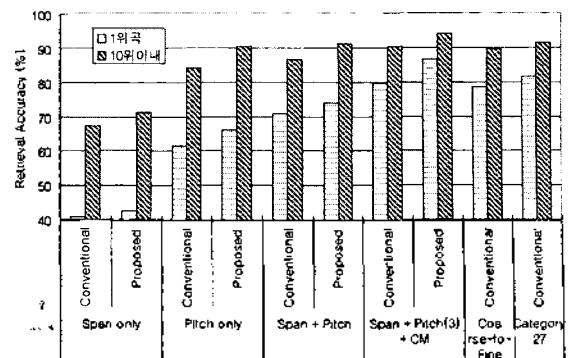


표 3. 다양한 특징량에 의한 검색정도의 비교
Table 3. Comparisons of the retrieval accuracy by various features.

78.4% 와 81.6%였다. 그러나 제안한 특징량을 사용한 경우 "IOI+ Pitch(3)+ CM"가 종래의 "Category 27"의 결과 보다 검색정도는 1위 랭크곡에 대해서 약 5%의 높은 검색 결과를 얻었다. 따라서 제안한 다이내믹 멜로디 표현과 복수피치후보 방법을 동시에 사용한 경우 종래의 방법보다 좋은 결과를 얻은 것을 알 수가 있다.

VIII. 결 론

사용자의 입력 허밍의 오류 및 신호 처리시 발생하는 오류 등에 강인한 악곡 검색 기법을 제안하였다. 실제 사용자의 입력 허밍은 개인차에 의한 템포, 장조의 변화와 애매한 기억에 의한 음표의 삽입/탈락 등의 여러 가지 요인으로 인해 데이터베이스 곡과 다른 입력이 들어 오게 된다. 이러한 문제점을 해결하기 위해서 최적의 상대값을 DP 패스에 따라 다이내믹하게 취하는 방법을 제안하였다. 한편 입력허밍이 완벽하다고 할지라도, 허밍으로부터 정확한 피치 정보를 얻기가 어렵다. 이러한 피치추출 오류를 고려하여 복수피치추출 방법을 제안하였다. 제안된 방법은 3개까지의 복수피치후보를 사용할 경우 99.7%의 검출 성능을 보였다. 또한 유사도 측정 알고리즘에 있어서 제안된 특징량을 효율적으로 사용하고 피치추출 오류에도 강한 매칭을 위해 DP평면의 공간을 3차원으로 확장하였다.

평가실험에서 제안된 방법의 경우 top-10의 검색정도는 94.1%의 결과를 얻었다. 이것은 종래 MIR 시스템보다 우수한 성능의 결과를 보였다.

참 고 문 헌

1. D. Feng, W.C.Siu, and H. Zhang, "Multimedia Information Retrieval and Management: Technological Fundamentals and Applications," Springer-Verlag, 2003.
2. Anni R. Coden, Eric W. Brown, and Savitha Srinivasan, "Information Retrieval Techniques for Speech Applications (LNCS 2273)," Springer-Verlag, 2002.
3. T. Kageyama, K. Mochizuki, and Y. Takashima, "Melody Retrieval with Humming," In Proc. Int. Computer Music Conference, 1993.
4. 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.
5. Tomonari Sonoda, Makataka Goto, and Yoichi Muraoka,

"WWW-based Music Retrieval System," pp.343-352, In Proc. ICMI'98, 1998.

6. A. Klapuri, "Pitch Estimation Using Multiple Independent Time-Frequency Windows," In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp.115-118, 1999.
7. Chai Wei, and Vercoe Barry, "Folk Music Classification Using Hidden Markov Models," In Proc. International Symposium on Music Information Retrieval, Oct. 2000.
8. S. Pauws, "CubyHum: A Fully Operational Query by Humming System," In Proc. ISMIR 2002, 2002.
9. 이지연, 한만수, "허밍 운율정보를 이용한 곡목 검색 기술," 대한 음성학회 학회지, 말소리 44, pp. 131-143, 2002.
10. 노승민, 아수철, 황인준, "대용량 오디오 데이터베이스에서의 효율적인 오디오 검색 기법," 데이터베이스 연구회지, 19(3), pp. 46-57, 2003.
11. 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 (Demo paper), pp. 636-637, 2001.
12. Naoko Kosugi, Yuichi Nishihara, Tetsuo Sakata, Masashi Yamamuro, and Kazuhiko Kushima, "A Practical Query-By-Humming System for a Large Music Database," ACM Multimedia 2000, pp. 333-342, 2000.
13. MiDiLiB, University of Bonn, Web pages (see <http://www-mmdb.iai.uni-bonn.de/forschungsprojekte/midilib/englisch/>).
14. R. J. McNab, L. A. Smith, I. H. Witten, C. L. Henderson, and S. J. Sunningham, "Toward the digital music library: tune retrieval from acoustic input," In Proc. ACM Digital Libraries, Bethesda, 1996.
15. Rodger J. McNab, Lloyd A. Smith, and Ian H. Witten, "Signal Processing for Melody Transcription," In Proc. of the 19th Australasian Computer Science Conference, 1996.
16. Sung-Phil Heo, Motoyuki Suzuki, Akinori Ito, Shozo Makino, and Hyun-Yeol Chung, "Error Tolerant Melody Matching Method in Music Information Retrieval," LNCS 3094, pp.212-227, Springer-Verlag 2004.
17. L. R. Rabiner, and B. H. Juang, "Fundamentals of Speech Recognition," Prentice-Hall, 1993.
18. Toshihiko Shirokaze, Shozo Makino and Ken'iti Kido, "Extraction of Fundamental Frequency Using Temporal Continuity over an Input Speech," Trans. IEICE 73-A(9) pp.1537-1539, 1990.
19. Sung-Phil Heo, Motoyuki Suzuki, Akinori Ito, and Shozo Makino, "Three Dimensional Continuous DP Algorithm for Multiple Pitch Candidates in Music Information Retrieval System," In Proc. ISMIR 2003, 2003.
20. J. Paulus and A. Klapuri, "Measuring the Similarity of Rhythmic Patterns," In Proc. ISMIR 2002, 2002.

저자 약력

• 허성필 (Sung-Phil Heo)



1989년: 영남대학교 전자공학과 (공학사)
 1993년: 영남대학교 대학원 전자공학과 (공학석사)
 2004년: 일본 동북대학교 대학원 정보과학연구과 (정보공학박사)
 1993년현재: KT 서비스개발연구소 선임연구원
 학회활동: 한국음악학회, ACM, 일본음악학회 각 회원

관심분야: 악곡정보 검색, 음성신호 처리, 저능망 서비스

• 정 현 열 (Hyun-Yeol Chung)



- 1975년: 영남대학교 전자공학과 졸업
- 1989년: 일본 동북대학교 대학원 정보공학과 수료(공학 박사)
- 1989년 현재: 영남대학교 전자정보공학부 교수
- 1992.7-1993.7: Carnegie-Mellon 대학 Robotics 연구소 객원연구원
- 1994.12-1995.2: 일본 토요하시기술과학대학 외국인 연구자
- 2000.6-2000.8: 미국 Qualcomm Inc. 수석 엔지니어

학회활동: 한국음향학회(부회장), IEEE, 한국통신학회, 일본음향학회, 일본 전자정보통신학회 각 회원

관심분야: 음성인식, 화자인식, 음성합성 및 DSP 응용분야