

가변 템포를 고려한 자동 음악 채보

Automatic Music Transcription Considering Time-Varying Tempo

주영호, 바니아 바부카지, 이준환
전북대학교 전자정보공학부 컴퓨터공학

Youngho Ju(nightmute@jbnu.ac.kr), Baniya Babukaji(everwith_7@jbnu.ac.kr),
Joonwhan Lee(chlee@jbnu.ac.kr)

요약

시간에 따라 변화하는 노래의 템포는 자동 음악 채보의 음길이 부호화에 있어 오류를 발생시키는 원인 중 하나이다. 본 논문에서는 변화하는 템포를 반영하여 음길이 부호화를 수행하는 개선된 자동 음악 채보 방법을 제안하였다. 제안된 방법에서는 단음 노래에서 휴지기를 이용하여 마디를 찾고, 마디의 연주시간, 즉 템포를 추정하였다. 추정된 템포는 발생된 개별 음길이를 조정하는데 활용하여 음길이 인식에 반영하였으며, 악보와의 일치도를 증가시켰다. 남성 및 여성이 단음으로 부른 16곡의 동요에 대한 실험결과 14 곡에서 정확하게 마디위치를 찾을 수 있었으며, 음길이는 약 89.4%, 음정은 약 84.8%의 원본 악보와의 일치도를 달성하였다.

■ 중심어 : | 자동음악채보 | 마디찾기 | 가변템포 |

Abstract

Time-varying tempo of a song is one of the error sources for the identification of a note duration in automatic music recognition. This paper proposes an improved music transcription scheme equipped with the identification of note duration considering the time-varying tempo. In the proposed scheme the measures are found at first and the tempo, the playing time of each measure, is then estimated. The tempo is then used for resizing each IOI(Inter Onset Interval) length and considered to identify the accurate note duration, which increases the degree of correspondence to the music piece. In the experiment the proposed scheme found the accurate measure position for 14 monophonic children songs out of 16 ones recorded by men and women. Also, it achieved about 89.4% and 84.8% of the degree of matching to the original music piece for identification of note duration and pitch, respectively.

■ keyword : | Automatic Music Transcription | Finding Measure Position | Time-varying Tempo of Song |

1. 서론

최근 음성 신호 처리기술이 발전하면서 자동 음악 채보 시스템을 비롯한 음악 감성 인식 및 검색 등에 대한 연구가 활발히 진행되고 있다[1-8]. 자동 음악 채보 시

스템은 기존의 음악에 익숙한 전문가가 직접 노래를 듣고 채보하는 방법과는 다르게, 노래가 가진 음악적 특징을 시스템이 자동으로 인식하여 악보로 만들어 주기 때문에 작곡의 보조적인 수단으로 활용할 수 있다. 또한 자동 음악 채보 시스템의 결과는 악보를 활용한 음

* 본 논문은 한국연구재단 일반연구자 지원사업에 의해 수행된 연구임.(NRF-2011-0022152)

접수번호 : #120910-002

접수일자 : 2012년 09월 10일

심사완료일 : 2012년 10월 05일

교신저자 : 이준환, e-mail : chlee@jbnu.ac.kr

원 검색, 악보기반 음악의 감성 분류 및 검색, 가창 교육 프로그램개발 등 다양하게 응용될 수 있다[9].

자동 악보채보 방법은 시간에 따라 변화하는 템포를 고려하지 않는 경우와 이를 고려하는 경우로 나누어 볼 수 있다. 만약 사람이 악보대로 기계적으로 노래한다거나 연주한다면, 자동채보에서 템포의 변화를 고려한다는 것은 박자의 변화가 없는 경우에는 무의미하다. 왜냐하면 동일한 길이의 음표는 동일한 시간구간 동안에 연주되며, 따라서 온셋(onset)정보를 알면 음표의 길이를 정확하게 예측해 낼 수 있기 때문이다.

그러나 사람들은 고의든 아니든 동일한 템포로 노래하거나 연주하지 않는다. 따라서 어떤 부분에서는 음악이 빨라지거나 또는 느려지며 동일음표를 연주한다 하더라도 어떤 부분에서는 음길이가 짧거나 길어 질수 밖에 없다. 실제 노래의 경우 부르는 사람에 따라 동일한 노래에서 템포가 약 1.43배 정도로 변화되는 경우도 발생하는 것을 관찰할 수 있었다. 이러한 시간에 따른 템포의 변화는 자동채보에 있어 음길이를 구분하는데 어려움을 제공한다.

박은중 및 형아영 등의 기존 자동 음악채보방법에서는 템포가 음악전체에 대해 일정하다고 가정하였다 [8][10]. 즉 박은중의 방법[10]에서는 음표의 조합규칙을 활용하여 첫 번째 마디의 위치를 인식하고 마디들이 동일한 지속시간을 가진다는 가정, 즉 동일 템포라는 가정 아래 연속적으로 마디를 나누었다. 이렇게 나누어진 마디 연주시간으로부터 4박자 계통 및 3박자 계통으로 박자를 추정하고, 추정된 박자에 따라 한 마디를 4등분 또는 3등분 하여 4분 음표의 길이를 가정하고, 온셋 사이의 시간간격 즉 *IOI*(Inter Onset Interval)에 따라 음길이를 부호화 하였다. 따라서 템포가 변화하게 되면 박자의 예측이 부정확해 질뿐만 아니라 부호화된 음길이 역시 큰 오차를 포함할 수 있다.

형아영의 방법[8]에서도 역시 템포가 일정하다는 가정 하에 *IOI*들의 군집화를 이용하였다. 따라서 이 방법에서도 템포가 변화하면 부정확한 군집화가 이루어 질 수밖에 없다. 즉 어떤 마디에서는 4분 음표가 8분 음표와 유사한 길이로 빠르게 발생되었다면 이는 8분 음표로 부호화될 수밖에 없으며, 한 마디의 음길이의 합이

동일하지 않는 경우, 즉 변박자라는 가정을 할 수밖에 없는 경우가 생길 수 있다.

본 논문에서는 자동 채보에 있어 변박자가 없다고 가정하고 변화하는 템포를 반영한 음길이 인식방법을 제안하였다. 제안된 방법에서는 휴지기 정보를 이용하여 마디를 추출하고, 이들 마디의 지속시간의 변화 즉 템포 변화를 *IOI*에 반영하여 음길이를 부호화하였다.

또한 제안된 방법에서는 형아영 등이 제안한 휴지기를 이용하는 마디 찾기 방법을 개선하였으며, 박은중 등이 제안한 박자결정 방법과 음길이 부호화 방법을 개선하였다.

자동 음악채보에서 채보의 대상이 되는 음악은 한 사람의 노래만을 대상으로 하는 단일 음(monophonic)의 경우와 여러 사람 혹은 반주가 가미된 복합 음(polyphonic)의 경우로 나누어진다. 본 논문에서 제안된 방법은 비트 히스토그램, 스펙트랄 에너지로부터의 온셋 에너지 함수, 리듬의 통계적 모형 등을 이용하는 방법들에서 대상으로 하는 악기로 연주되는 복합음이 아니라 녹취된 단일음의 노래를 대상으로 한다[11][12]. 따라서 제안된 방법은 템포의 변화가 심한 경우, 예를 들어 비전문가가 흥얼거리는 허밍을 악보로 만들어 작곡하는데 도움을 줄 수도 있으며, 악보의 유사성을 이용한 내용기반 음악검색의 전처리기로 사용될 수 있다.

본 논문의 II장에서는 가변템포를 반영한 단음 음악채보를 위한 음정과 *IOI* 추출, 마디검출 및 박자인식, 음길이 및 음정 부호화 방법을 서술하고, III장에서는 실험 및 결과, IV장에서는 결론을 제시하였다.

II. 가변 템포를 반영한 단음의 음악채보

1. 음정추출

음악의 최소 단위인 개별 음들을 찾아내는 일은 자동 음악 채보의 성능을 좌우하는 기본적인 작업 중 하나이다. 따라서 개별 음절을 찾는데 기초가 되는 음정(pitch)의 추출은 자동채보의 정확성을 위해 대단히 중요하다. 본 논문에서는 정확한 음정추출을 위해 Praat 프로그램을 이용하였다. [그림 1]은 채보를 위해 녹음된

음원(.wav)파일에서 Praat 프로그램을 이용해 음정성분을 추출하는 모습을 보여주고 있다[13].

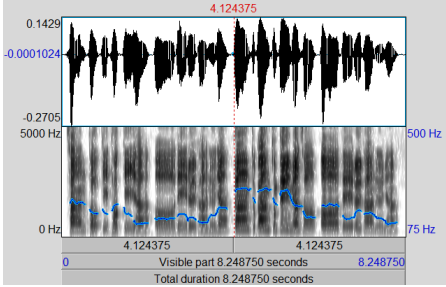


그림 1. Praat 프로그램을 이용한 동요 '산토끼' 음정성분 추출

2. BIC를 이용한 IOI(Inter Onset Interval) 추출

개별 음 분리를 위해서는 추출된 음정 데이터에서 음성구간과 비음성(pause) 구간을 구분하고, 음성구간에서는 음정이 동일하게 유지되는 IOI들을 추출해야한다. 제안된 방법에서는 이를 위해 BIC (Bayesian Information Criteria)알고리즘을 이용하였다[14]. 이 방법은 형아영 등의 SIDE(Stabilized Inverse Diffusion Equation)의 경우[10]와 같이 반복횟수 및 정지조건 등을 수동으로 조절할 필요가 없으며, 박은중 등의 GMM(Gaussian Mixture Model)에서 나타나는 시간축 상의 분할의 불일치 현상도 일어나지 않으며, 문턱치 등을 조절할 필요 없이 BIC 측면에서 최적의 분할 결과를 제공해준다.

데이터 집합 $\{x_1, x_2, \dots, x_N\}$ 을 분할할 경우 최대 우도(maximum likelihood) 통계는

$$R(i) = N \cdot \log|\Sigma| - N_1 \cdot \log|\Sigma_1| - N_2 \cdot \log|\Sigma_2| \quad (1)$$

와 같이 정의 된다. (식 1)에서 N_1 은 첫 번째 세그먼트의 데이터 수이며, N_2 는 두 번째 세그먼트의 데이터 수이다. 또한 Σ , Σ_1 , Σ_2 는 각각 전체, 첫 번째 세그먼트, 두 번째 세그먼트의 공분산 행렬이다. 분할 위치 i 에서 BIC 는

$$BIC(i) = R(i) - \lambda P \quad (2)$$

와 같으며, (식 2)에서 P 는 패널티 항으로

$$P = \frac{1}{2} \left(d + \frac{1}{2} d(d+1) \right) \log N \quad (3)$$

과 같다. (식 2)와 (식 3)에서 $\lambda = 1$ 이며 d 는 데이터의 차원으로 피치 데이터의 경우 1차원이다. 최적의 분할 경계는 최대우도추정에 의해 (식 4)와 같이 결정할 수 있다.

$$\hat{t} = \arg \max_i BIC(i). \quad (4)$$

만약 여러 개의 음절분할이 요구되는 경우에는 [표 1]과 같은 순차적인 알고리즘을 적용할 수 있다. [그림 2]는 BIC를 이용한 음절 분할 결과를 보여주고 있다.

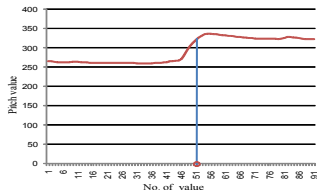
표 1. BIC 기반 음절 경계 탐색 알고리즘

step 1	구간 초기화 $[a, b] : a = 1; b = 2$
step 2	BIC를 이용하여 $[a, b]$ 사이에 변화하는 점이 있는가를 검출
step 3	만약 변화 지점이 $[a, b]$ 사이에 없으면 $b = b + 1;$ 그렇지 않고 \hat{t} 가 검출된 변화점이라면 $a = \hat{t} + 1; b = a + 1;$ 알고리즘 재시작;
step 4	step 2 를 반복

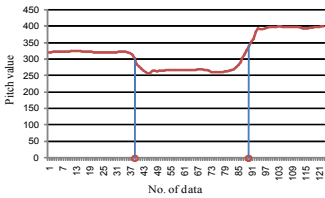
제안된 방법에서는 [표 1]의 과정을 통해 분할된 음절을 개별 음으로 보고, 개별 음이 유지되는 기간을 IOI, 즉 동일음정 구간으로 간주한다. 그러나 노래 부르는 사람의 음정이 불안한 경우에는 악보와 다르게 음절 분할이 파편화되어 더 많은 IOI들이 추출되는 경향이 있다. 따라서 이러한 파편화에 따른 음정채보 불안정함을 완화하기 위해 인접한 IOI의 피치 값이 유사할 경우에는 연속된 IOI를 단일 IOI로 병합하였다. 인간이 인지하는 음정의 차이는 피치의 로그 값에 관계되기 때문에 (식 5)를 만족하는 경우 두 IOI를 병합하였다.

12음계에서 한 음과 인접음의 차이는 피치의 로그함수를 취할 경우 0.05이므로 임계값은 0.025로 설정하였다.

$$\log(IOI_{n-1} \text{ 평균피치}) - \log(IOI_n \text{ 평균피치}) < th. \quad (5)$$



(a) 두 개로 분할되는 경우



(b) 세 개로 분할되는 경우

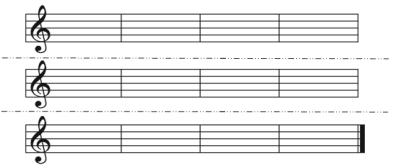
그림 2. BIC 방법에 의해 분할된 음절

3. 마디 검출 및 가변템포 반영 방법

제안된 방법의 마디 검출 단계에서는 음정 추출 단계에서 찾을 수 있는 비음성 구간정보를 이용한다. 일반적으로 비음성 구간 즉 휴지기간의 끝은 악보 상에서 마디의 후보 위치라 가정할 수 있다. 또한 노래의 중간, 특히 소절(phrase)사이에는 긴 휴지기가 존재하게 되는데 제안된 방법에서는 이런 휴지기 정보들을 이용하여 마디를 검출하게 된다.



(a) 2 소절단위



(b) 3 소절 단위

그림 3. 마디 찾기를 위한 소절 위치 찾기

[그림 3]의 (a)에서 볼 수 있듯이 2ⁿ개의 소절을 가지는 2소절 타입 음악의 경우는 노래의 1/2 위치에 긴 휴지기가 나타나며, [그림 3]의 (b)경우와 같은 3ⁿ개의 소절을 가지는 3소절 타입의 음악의 경우 1/3 지점과 2/3 지점에 긴 휴지기가 나타난다. 즉 (식 6)과 같이 긴 휴지기(Max Pause)들이 곡 전체상에서의 어떤 위치에 있는가에 따라 어떤 소절 타입의 노래인가를 알아낼 수 있으며, 소절을 분리해 낼 수 있다. 본 논문에서는 2소절 및 3소절 타입만을 고려했는데 5소절, 7소절 타입 등의 구분도 동일한 방법으로 가능하다.

$$\begin{aligned} & \text{If } (1/2 - \Delta_2 < \text{Max Pause} < 1/2 + \Delta_2) \quad (6) \\ & \quad \text{then 2 Phrase Type} \\ & \text{Elseif } (1/3 - \Delta_3 < \text{Max Pause} < 1/3 + \Delta_3) \\ & \quad \text{AND } (2/3 - \Delta_3 < \text{Max Pause} < 2/3 + \Delta_3) \\ & \quad \text{then 3 Phrase Type} \\ & \text{Elseif} \end{aligned}$$

일단 소절의 위치를 찾아내면 한 소절에 2^m개의 마디가 존재하기 때문에, 소절 내에서의 마디구분은 본 논문에서 제안하는 ‘이진 구간 분리법’을 이용하여 수행할 수 있다. ‘이진 구간 분리법’에서는 휴지기를 중심으로 또는 휴지기가 없는 경우에는 가장 가까운 IOI의 끝지점을 기준으로 마디분할을 재귀적으로 수행한다.

표 2. 이진구간 분리 알고리즘

step 1	휴지기 끝점 집합 = E ; 마디 수 = 1, 마디 길이 = T ;
step 2	현 수준의 모든 마디 중간위치를 E 에서 탐색 만약 E 에 존재하면 탐색된 위치로 마디분할 그렇지 않으면 가장 가까운 IOI 끝점으로 마디 분할
step 3	모든 마디분할이 IOI 끝점에서라면, 이진분할 중단; 이진 수준의 마디분할 결과를 출력; 그렇지 않다면 step2 반복;

한 소절 내에서 [표 2]를 이용한 마디 위치 결정방법에서는 먼저 주어진 소절의 휴지기들을 찾아낸다. 이들 휴지기 즉 비음성 구간의 끝부분의 시간집합을 E 라 하고 전체 소절의 길이를 T 라 하자. 먼저 $T/2$ 부분에서 충분히 가까운 휴지기의 끝을 집합 E 에서 찾고, 이 휴지기의 끝을 기준으로 소절을 두 개의 마디로 분할한

다. 다음 수준(level)에서 이 마디들은 같은 방법으로 마디를 반으로 나눈 시점에서 충분히 가까운 휴지기 끝을 집합 E 에서 탐색하며 재귀적인 마디분할을 시도한다. 만약 이 과정에서 집합 E 에 충분히 가까운 휴지기가 존재하지 않으면, 가장 가까운 IOI 의 끝점을 마디위치로 간주한다. 이와 같이 ‘이진 구간 분리법’에서는 재귀적인 과정을 반복하며 마디 찾기를 수행하며, 만약 어떤 수준에서 마디위치를 집합 E 에서 하나도 찾지 못할 경우 탐색을 중단하며, 이전 수준에서 결정된 마디를 최종적인 마디 탐색결과로 제시한다.

제안된 방법은 휴지기 정보를 이용한다는 점에서 형아영 등의 방법과 유사하다. 그러나 내용적으로는 구간을 반으로 분할하여 휴지기 또는 IOI 의 끝점을 탐색하고 이를 발견하면 이를 분할 기준점으로 하고 분할을 계속적으로 진행해 나가는 재귀적인 방법을 적용하는 부분에서 형아영의 방법과 구별된다[10]. 즉 형아영 등의 방법에서는 음악 전체의 템포가 변하지 않는 점에서 출발하여 순차적으로 휴지기 정보를 탐색하였지만, 본 논문에서는 음악이 지속되는 동안에 템포가 연속적으로 변화한다는 점을 염두에 두고 재귀적인 이진탐색을 진행한다.

마디분할 과정을 거치면 각각의 마디의 길이 즉 마디의 지속시간을 알 수 있으며, 이 마디의 지속시간은 시간에 따라 변화되는 템포정보를 포함하고 있다. 즉 박자의 변화가 없다는 가정 아래 마디길이가 긴 경우에는 템포가 느린 구간이며, 템포가 빠른 경우에는 마디길이가 짧아진다.

이러한 템포의 변화에 따라 동일한 음표라도 발생되는 기간이 달라지기 때문에 템포의 변화를 고려하지 않으면 음길이 부호화의 오류를 포함 할 수밖에 없으며, 템포변화에 무관한 음길이 부호화를 위해서는 템포정보를 반영하여 음길이를 정규화 할 필요가 있다.

본 논문에서는 모든 마디의 길이의 평균값 D_{ave} 를 구하고 이를 기준으로

$$IOI'_i = \frac{D_{ave} \cdot IOI_i}{D} \quad (7)$$

과 같이 음길이들을 정규화 하였다. (식 7)에 따르면 조

정된 IOI'_i 는 음절 분할된 IOI_i 에 비례하고, 음절이 속한 마디의 길이 D 에 반비례한다. 즉 평균마디 길이 D_{ave} 보다 긴 마디에 속하는 IOI_i 의 길이는 짧아지고 그 반대의 경우는 길어지며, 조정된 IOI'_i 들을 포함하는 모든 마디길이들은 평균값 D_{ave} 로 동일하게 된다. 즉 제안된 방법에서는 템포변화를 가정하지 않은 IOI_i 대신에 조정된 IOI'_i 들이 사용하여 템포에 무관한 박자 검출 및 음길이 인식을 수행한다는 점에서 기존의 방법과 차별화 된다.

4. 박자 검출방법

4분음을 기준으로 하는 노래에서 박자는 한 마디 안에 4분 음표(♩)가 몇 개 들어가느냐에 따라 결정될 수 있다. 4분 음표가 마디 안에 3개가 들어 있다면 3박자 계통이고 4개가 들어 있다면 4박자 계통이라고 판단할 수 있다. 본 논문에서는 3/4박자, 6/8박자는 3박자 계통으로, 2/4박자와 4/4 박자는 4박자 계통으로 구분하였다.

제안된 방법에서는 가설검증을 바탕으로 4박자 계통과 3박자 계통을 구분하고 4박자 계통의 경우는 부호화되는 최소의 음표가 16분 음표라는 가정을 토대로 2/4와 4/4 박자를 구분하였다. 제안된 방법에서는 정규화된 한 마디를 4등분하고 이 길이를 4분 음표로 가정했을 경우(4박자 가설)와, 3등분하고 이 길이를 4분 음표로 가정했을 경우(3박자 가설)를 비교하여 4박자 계통과 3박자 계통을 구분하였다. 박자를 찾는 알고리즘은 [표 3]과 같다.

표 3. 가설검증을 통한 박자 찾기 알고리즘

step 1	정규화된 마디를 이용한 4박자 가설 설정 4로 나누고 이 길이를 4분 음표로 가정했을 경우
step 2	4박자 가설을 기반으로 16분 음표로부터 온음표까지의 길이 결정; 4박자가설 = 0; 3박자가설 = 0;
step 3	각각 마디의 IOI'_i 들을 가장 가까운 길이의 음표로 부호화; 부호화된 마디의 음표의 합 > 3.5 이면 4박자가설 ++; 아니면 3박자 가설++;
step 4	4박자 가설 >= 3박자 가설 이면 4박자 계통 그렇지 않으면 3박자 계통;

[표 3]의 박자 찾기 알고리즘에서는 일단 노래가 4박자라는 가정을 토대로 정규화된 마디길이 D_{ave} 를 4로 나누어 이를 4분음표(♩)의 길이로 간주하며, 이 길이에 0.25(♩), 0.5(♩), 0.75(♩), 1(♩), 1.5(♩), 2(♩), 3(♩), 4(♩)를 곱하여 16분 음표로부터 온음표까지의 길이를 결정한다. 이러한 가정을 기반으로 결정된 음표 길이로 각 마디의 조정된 IOI_i 즉 음절들을 임시 부호화하고, 이들의 합이 3.5 이상이면 4박자 가설에 그렇지 않으면 3박자 가설에 투표표를 진행한다. 이런 과정을 통해 모든 마디에 대한 가설의 투표가 진행되며 최종적으로 더 많은 득표를 한 가설이 채택된다.

또한 제안된 방법에서 2/4박자와 4/4박자의 구별은 [표 3] step 3의 4/4박자를 가정한 임시 부호화 과정에서 노래 전체에서 16분 음표가 존재하지 않으면 2/4박자로 간주하였다. 이는 만약 16분 음표가 임시부화 과정에서 나타났다면 2/4박자에서는 32분 음표에 해당되어 최소 음이 16분 음표보다 작아지게 되기 때문이다. 이러한 구별은 원래 2/4박자 악보에 32분 음표가 존재하지 않는 경우에는 정확한 판별결과를 제공할 수 있으나, 그렇지 않는 경우에는 2/4박자로 구별할 수 없다.

제안된 방법은 박자의 변화가 없다는 가정을 바탕으로 하고 있다. 만약 박자의 변화, 즉 변박이 가능하다고 가정하면 투표과정 없이 각 마디에서 음표길이의 합이 3.5 이상이면 4박자 계통, 그렇지 않으면 3박자 계통으로 결정할 수도 있다.

5. 음길이의 부호화 및 보정

제안된 방법의 박자 결정이 끝나면 4분 음표(♩)의 지속시간을 알아낼 수 있다. 즉 4/4박자의 경우는 평균 마디길이를 4등분한 것이며, 2/4박자의 경우는 2등분한 길이, 또한 3박자 계통의 경우는 3등분한 길이에 해당된다.

따라서 이 길이에 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4를 곱하면 음길이 부호화에 사용될 표준길이 산출되며, 조정된 IOI' 또는 휴지기들의 길이를 이들 표준길이와 비교하는 방식으로 음표부호화가 진행된다.

제안된 방법에서는 16분 음표 이하의 음은 고려하지

않으므로, IOI' 가 최소 음길이 즉 16분 음표 보다 짧을 경우, 즉 BIC에 의한 음절분할 과정에서 파편화에 의해 발생한 짧은 IOI' 는 이전 IOI' 와 병합하여 부호화 하였다.

이러한 부호화 과정을 통해 IOI' 또는 휴지기들을 음표로 옮기면 앞서 언급한 바와 같이 음악이 3박자 계통이면 부호화된 음표길이의 총 합이 3에 가까워지고 4박자 계통이면 부호화된 값의 총 합이 4 또는 2에 가까워지게 된다. 그러나 이 음표부호화 결과 한 마디내의 음표의 합이 박자에 따라 4, 3, 2 와 같이 정확하게 맞아 떨어지지 않는 경우도 발생한다. 즉 음표로 옮기는 양자화 과정에서의 근사화로 발생한 오차들의 합이 0이 되지 않는 경우들이 발생한다는 것이다. 따라서 음표의 부호화를 끝내고 난 뒤에 음표길이의 합이 마디를 다 채우지 못하거나 남는 경우에는 음표 길이 보정이 필요하다.

제안된 방법에서는 음절의 길이 IOI' 와 음표의 표준 길이사이의 오차를 고려하여 음표의 길이를 조정하였다. [표 4]는 음표보정 알고리즘을 보여주고 있다.

표 4. 음표보정 알고리즘

step 1	보정대상 마디의 음표에 대해 오차 = 해당음표의 IOI' - 표준 음표길이
step 2	음표의 합이 박자 수에 모자랄 경우 오차가 가장 큰 IOI' 에 대해 상위음표로 보정; 음표의 합이 박자 수를 넘을 경우 오차가 가장 작은 IOI' 에 대해 하위음표로 보정;
step 3	음표의 합이 박자수와 같지 않을 경우 step 2를 반복;

[표 4] step 2의 음표보정 과정에서 부호화된 음표의 합이 박자 수에 모자랄 경우에는 음표의 길이를 늘려야 하는데 오차가 가장 큰, 즉 부호화된 표준음 보다 실제 길이가 가장 긴 IOI' 의 음표를 한 단계 늘려준다. 반면에 그 반대의 경우에는 음의 오차 값을 가지는 IOI' 중에 가장 작은 오차를 가지는 (절대치로 볼 때는 가장 큰 오차) IOI' 를 찾아 한 단계 줄여준다. 이러한 보정에 의해 경우에 따라서는 박자수를 넘기도 하고 부족하기도 하기 때문에 음표길이의 합이 박자수와 같을 때 까지

이 과정을 반복한다.

실제 구현과정에서는 보정대상이 긴 길이의 음일 경우, 보정하고 난 후에 길이 변화폭이 커지기 때문에 보정대상은 4분 음표 이하의 음길이로 제한하였다. 즉 16분 음표, 8분 음표, 점 8분 음표, 4분 음표를 보정대상으로 하였다.

6. 음정 부호화 방법

미국 표준 협회표준(American National Standards Institute, 1960)은 음정을 “소리의 높낮이 척도 속에서 어떤 높이에 해당하는지 판단하는 청각적 감각적 특성”이라고 정의하고 있다[15]. 본 논문의 주요 관점은 템포변화 고려한 음길이 부호화이며 음정의 부호화와는 큰 관련성이 없다. 그러나 음길이의 변화가 음정의 부호화 정확성에도 영향을 미치며 전적으로 단일음의 자동채보 성능에도 영향을 미치기 때문에 음정부호화 방법을 약술한다.

음정 부호화 방법에는 절대음정을 이용하는 방법과 상대음정을 이용하는 방법이 있다[16]. 절대음정을 이용할 경우에는 노래의 건(key)을 고려하여 부호화할 수 있으며, 상대음정일 경우에는 건 정보는 무시되고 모든 음이 다장조를 기준으로 부호화 된다. 본 논문에서는 형아영 등이 제안한 음정부호화 방법을 활용하였다. 이들은 사람이 음정을 인지할 때와 같이 외부에서 주어지는 기준 음에 의존하여 음정을 판별하는 상대음정 부호화 방법을 사용하였다. 즉 앞 음보다 얼마나 변화했는지를 측정하고, 그 변화 정도를 음정 부호화에 사용할 12음계에 적용할 수 있도록 근사하고, 이를 이용하여 음정부호화를 시도하였다.

제안된 방법에서 음정부호화 과정은 이러한 형아영 등의 상대음정 부호화 방법을 이용하여 각 *IOI*'의 대표 음정을 결정하는 과정으로 수행되었으며, 노래하는 사람의 음정불안의 영향을 최소화하기 위해 외톨이(outlier) 데이터를 배제한 알파 절삭 평균(trimmed mean)을 이용하였다. 실험에서 알파 값은 20%로 하였다.

III. 실험 및 결과

본 논문의 실험에서 사용된 음원은 14곡의 동요음원과 2곡의 애국가 음원으로 총 16개의 음원이었으며, 악기소리 없이 오직 여성과 남성의 목소리만 녹음되었다. 이들 음원은 개인용 컴퓨터 앞에서 마이크를 통하여 수집하였고, praat 소프트웨어를 이용하여 각 음원의 음정을 추출하였다.

[그림 4]는 본 논문에서 개발한 자동채보 시스템의 사용자 인터페이스의 모습을 보여주고 있다. 개발된 인터페이스에서는 [그림 4]의 명령 버튼들을 이용하여 채보단계별로 praat으로부터 얻은 피치데이터를 읽기, BIC를 이용한 음절분할, 소절 및 마디 추출, 음길이 부호화 및 음정 부호화 등의 명령을 지시할 수 있으며, 최종의 채보결과는 우측의 창에 텍스트 형식으로 표현된다. 따라서 [그림 6]과 [그림 7]의 실험결과는 텍스트 형태의 채보결과를 nwc(noteworthy composer) 프로그램의 악보그리기를 활용하여 오선지에 옮겨 그린 것이다.

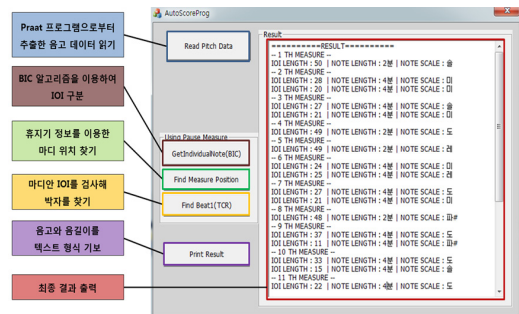


그림 4. 자동채보 시스템의 사용자 인터페이스

[표 5]에는 제안된 방법으로 각각의 음원에서 마디위치를 탐색하고 마디지속 시간의 변화, 즉 템포의 변화를 측정된 결과를 보여주고 있다. 서론에서 언급한 바와 같이 노래마다 마디 지속시간의 변화가 상당함을 알 수 있으며, 음원 song11은 가장 변화가 심한 경우로 전체적으로 약 1.43배의 템포변화가 있음을 알 수 있다. [그림 5]는 음원 song11의 시간에 따른 마디 지속시간의 변화를 보여주고 있다.

표 5. 노래의 마디 지속시간 (단위: 10ms)

음원 번호	최단 마디지속 시간	최장 마디지속 시간	평균 마디지속 시간
song 1	37	55	49
song 2	80	125	101
song 3	107	122	113
song 4	287	308	298
song 5	280	309	298
song 6	138	160	149
song 7	141	158	149
song 8	252	276	267
song 9	105	150	130
song 10	194	209	199
song 11	95	136	112
song 12	198	216	206
song 13	197	221	208
song 14	192	215	201
song 15	291	309	299
song 16	81	90	86

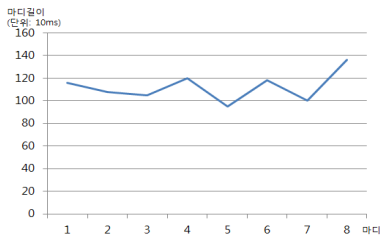


그림 5. 음원 song 11의 마디 지속시간의 변화

본 논문에서 제안된 방법에 따른 마디위치, 음길이와 음정인식의 실험결과는 [표 6]과 같다. 마디 찾기의 경우 제안된 방법은 16곡 중 14곡에서 정확한 마디를 찾아 주었다. 마디 찾기에 실패한 음원 song2의 경우는 휴지기 없이 이어져 노래되었기 때문에 원 악보보다 마디가 한 수준 덜 나누어진 경우이며, 음원 song16의 경우는 노래를 끊어서 불러서 마디가 한 수준 더 나누어진 경우에 해당된다.([표 2]의 step2 단계 참조) 제안된 방법에 의해 음원 song2의 탐색된 마디는 [그림 6]의 (a), 해당 원 악보는 [그림 6]의 (b)와 같으며, 음원 song16의 탐색된 마디는 [그림 6]의 (c), 해당 원 악보는 [그림 6]의 (d)와 같다. 이와 같이 마디를 덜 나누거나 더 나누는 경우는 박자추정의 오류 때문에 음표의 정확도를 원본 악보와 비교해 판별할 수 없다. 왜냐하면 [그림 5]에서 확인할 수 있듯이 인식된 음표의 길이가 원본의 2배 또는 1/2배에 해당되어 부호화되기 때문이다.

표 6. 원 악보와 비교하였을 때 정확도

곡번호	마디위치	음길이(%)	음정(%)
song 1	정확함	94	81
song 2	-	-	92
song 3	정확함	93	90
song 4	정확함	91	86
song 5	정확함	92	82
song 6	정확함	87	80
song 7	정확함	92	81
song 8	정확함	90	87
song 9	정확함	89	94
song 10	정확함	82	84
song 11	정확함	83	80
song 12	정확함	94	81
song 13	정확함	90	90
song 14	정확함	85	87
song 15	정확함	90	77
song 16	-	-	86

(a) 음원 song2의 채보결과

(b) 음원 song2의 원곡 악보

(c) 음원 song16의 채보결과

(d) 음원 song16의 원곡 악보

그림 6. 마디를 정확히 찾지 못한 채보결과

이러한 마디 찾기 실험결과는 박은종의 실험결과보다 우수하다[8]. 실제 박은종의 경우 동일한 음원 데이터에 대해서 첫 소절부분만을 고려하였음에도 3개의 노래에서 정확한 마디 추출에 실패하였다. [표 5]에서 확인된 바와 같이 실제로는 노래의 템포변화가 크기 때문에, 이를 고려하지 않고 첫 소절부분만을 실험대상으로 한 박은종의 결과는 노래의 전체를 대상으로 할 경우 더 많은 오류를 포함할 것으로 예상된다.

본 실험에서 채보된 결과를 원본 악보와 비교했을 경우 음길이의 경우, 약 89.4%의 정확도를 음정은 약 84.8%의 정확도를 보였다. 음길이의 부호화 결과에서 같은 음이 반복되어 하나로 합쳐지는 경우는 전체 음길이의 합이 맞을 경우 옳은 경우라고 간주하였다. 음정의 경우도 대부분 반음차이 정도만 존재하고 정확한 음을 제시하였다.

(a) 박은종 방법의 결과

(b) 제안한 방법의 결과

(c) 원곡 악보

그림 7. 음원 song12(학교중)에 대한 자동 채보결과

[그림 7]은 음원 song12(학교중)의 채보 결과로 [그림 7]의 (a)는 박은종의 제안한 방법의 채보 결과이며, [그림 7]의 (b)는 제안된 방법의 결과이다. [그림 7] (c)의 원본 악보와 비교를 통해 확인 할 수 있듯이 [그림 7]의 (a)의 박은종 방법의 경우 2 소절부터는 마지막 마디를 인식 하지 못하고 불안정하게 마무리 된 모습을 보여 주었다. 또한 첫 소절의 셋째 마디에서는 총 4박자에 16분 음표가 모자라게 채보 되었다. 반면 [그림 7]의 (b)에서 볼 수 있듯이 제안된 방법의 채보 결과는 실제 원본 악보 [그림 7]의 (c)와 비교 하였을 때, 마디위치가 완벽하게 일치함을 알 수 있다. 제안된 방법의 음정 인식 결과를 관찰하면 [그림 7]의 (b)에서 제 1소절의 두 번째 마디에서는 반음 내리고, 제 2 소절의 둘째마디와 셋째마디에서는 반음 오르내리는 음정인식의 오류를 보여주고 있다.

또한 제안된 방법의 결과인 [그림 7]의 (b)에서 알 수 있듯이 제 1소절에서 첫 번째 마디와 셋째 마디 제 2소

절의 첫째마디와 둘째 마디의 음길이 부호화가 원본 악보와 일치하지 않으나, 연속된 동일 음정이 (식 5)에 의해 병합되었기 때문에 나타난 현상으로 [표 6]의 실험 결과에는 올바른 인식결과로 간주하였다.

IV. 결론

본 논문에서는 사람이 부르는 노래와 같은 단일음원으로 부터의 자동채보에 있어 변화하는 템포를 반영한 음길이 부호화 방법을 제안하였다. 제안된 방법에서는 사람이 발성한 개별 음들의 정확한 분할 위치를 알아내기 위해 BIC를 활용하였으며, 비음성 구간 즉 휴지기 정보를 이용하여 소절과 마디위치를 찾았다. 이들 탐색된 마디의 지속시간의 변화, 즉 템포 변화는 개별 발생 음절의 길이 즉 IOI 에 반영되고 조정되어 음길이를 부호화하는데 이용하였다.

이러한 방법은 기존의 템포변화를 반영하고 있지 않았던 방법에서 문제시되었던 음길이 부호화의 부정확성을 극복하는 데 기여할 수 있으며 궁극적으로 자동채보의 정확성을 높일 수 있었다. 제안된 방법은 템포의 변화에도 불구하고 16개의 동요 음원에 대해 14곡에서 정확한 마디를 탐색해 주었으며, 음길이 부호의 정확도 약 89.4%, 음정 부호화의 정확도 약 84.8%의 결과를 얻을 수 있었다.

제안된 방법은 구전되는 노래의 자동채보, 악보를 기반으로 하는 내용기반 음원검색 등에 활용될 수 있을 것으로 기대된다. 현재 구현된 프로그램은 마디정보, 음길이와 음정의 채보결과를 텍스트 형식으로 출력하고 있으나, .nwc 자동파일변환 프로그램 등을 부착한다면 작곡보조 시스템과 연계할 수 있다[17]. 향후 국악의 자동 채보와 동요 이외의 다른 장르의 복합음원의 채보에 대해서도 지속적인 연구를 수행할 예정이다.

참고 문헌

- [1] Matija Marolt and Marko Privosnik, "SONIC : A System for Transcription of Piano Music,"

Cybernetics and Systems, Vol.33, No.6, pp.603-627, 2002.

[2] A. Martin, *Blackboard System for Automatic Transcription of Simple Polyphonic Music*, MIT Media Laboratory Perceptual Computing Section Technical Report No.385, 1996.

[3] Lloyd A. Smith and H. Ian, "Signal Processing for Melody Transcription," 19th Australian Computer Science Communications, pp.301-307, 1996.

[4] Anssi. P. Klapuri, "Automatic Transcription of Music," Stockholm Music Acoustics Conference, Institute of Signal Processing, 2003.

[5] Marolt, "A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music," IEEE Transactions on Multimedia, Vol.6, No.3, pp.439-449, 2004.

[6] 장준영, *퍼지적분을 이용한 곡조 인식 시스템의 설계와 구현*, KAIST 석사학위논문, 1996.

[7] 형아영, 이준환, "SIDE를 이용한 자동 음악채보 시스템," 한국정보처리학회 논문지, 제16권, 제2호, 2009.

[8] 박은중, 신송이, 이준환, "단일 음원 노래에서 음표의 조합 규칙을 이용한 마디 위치 찾기," 한국콘텐츠학회, 제9권, 제10호, pp.1-12, 2009.

[9] 지정규, *오디오 데이터베이스의 효율적 검색을 위한 선율질의 처리기*, 숭실대학교 박사학위 논문, 1998.

[10] 박은중, 신송이, 이준환, "단일 음원 노래에서 음표의 조합 규칙을 이용한 마디 위치 찾기," 한국콘텐츠학회, 제9권, 제10호, pp.1-12, 2009.

[11] Geoffroy Peeters, "Time Variable Tempo Detction and Beat Marking," International Computer Music Association, Barcelona, Spain, 2005.

[12] Kenichi Miyamoto, Hirokazu Kameoka, Haruto Takeda, Takuya Nishimoto, and Shigeki Sagayama, "Probailistic Approach to Automatic

Music Transcription from Audio Signals," Proceedings of IEEE Conf. on ASSP, Vol.2, pp.II-697-700, April 2007.

[13] Praat URL : <http://www.praat.org>.

[14] Scott Shaobing Chen and P. S. Gopalakrishnan, Speaker, *Environment and Channel Change Detection and Clustering via The Bayesian Information Criterion*, DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[15] 이석원, *음악음향학*, 심설당, 2003.

[16] 한국음악지각인지학회, *음악의 지각과 인지*, 음악세계, 2005.

[17] NoteWorthy Composer URL: <http://www.noteworthysoftware.com>.

저 자 소 개

주 영 호(Youngho Ju)

준회원



- 2010년 8월 : 전북대학교 전자정보공학부 컴퓨터공학과(공학사)
- 2011년 3월 ~ 현재 : 전북대학교 컴퓨터공학과 석사 과정

<관심분야> : 음성처리, 멀티미디어

바니아 바부카지(Baniya Babukaji)

정회원



- 2005년 11월 : Department of Computer Science, Pokhara University(공학사)
- 2010년 8월 : 전북대학교 전자정보공학부 전자공학과(공학석사)
- 2011년 9월 ~ 현재 : 전북대학교 컴퓨터공학과 박사 과정

<관심분야> : 음성처리, 멀티미디어

이 준 환(Joonwhoan Lee)

정회원



- 1980년 : 한양대학교 전자공학과 (공학사)
- 1982년 : 한국과학기술원 전자공학과(공학석사)
- 1990년 : 미국 미주리대학 전기 및 컴퓨터공학과(공학박사)
- 1990년 ~ 현재 : 전북대학교 전자정보공학부 교수
<관심분야> : 영상처리, 컴퓨터 비전, 인공지능