

## TV-필터와 최적 음표조합을 이용한 개선된 가변템포 음악채보방법

### An Improved Automatic Music Transcription Method Using TV-Filter and Optimal Note Combination

주영호\* · 이준환\*

Young-Ho Ju\*, Joonwhoan Lee\*\*

\*전북대학교 전자정보공학부 컴퓨터공학전공

† Division of Computer Science & Engineering, Chonbuk National University

#### 요 약

본 논문에서는 가변템포를 반영한 단일음악 채보의 정확성을 증가시키기 위한 기존의 방법을 개선하는 세 가지 방안을 제시하였다. 첫째는 TV 필터를 활용한 음정 데이터의 평활화로 음정분할 결과의 파편화 현상이 줄어들게 하였다. 또한 음정과 에너지, 규칙기반 방법을 융합한 마디탐색 방법으로 마디 탐색결과의 안정성을 향상시켰다. 뿐만 아니라 마디 내에서 음표의 합이 박자수와 같으면서 양자화 오차의 합을 최소화하는 최적의 방법으로 임시 음길이 부호화 결과를 보정하였다. 그 결과 16개의 동요 음원에서 완벽한 마디위치를 탐색하였으며, 음길이 부호의 정확도 약 91.3%, 음정 부호화 정확도는 약 86.7%의 개선된 결과를 얻을 수 있었다.

**키워드** : 자동음악채보, 단일음원, 가변템포, TV 필터, 음정 평활화, 마디탐색, 음길이 최적부호화

#### Abstract

This paper proposes three methods for improving the accuracy of auto-music transcription considering with time-varying tempo from monophonic sound. The first one that uses TV(Total Variation) filter for smoothing the pitch data reduces the fragmentation in the pitch segmentation result. Also, the measure finding method that combines three different ways based on pitch and energy of sound data, respectively as well as based on rules produces more stable result. In addition the temporal result of note-length encoding is corrected in optimal way that the resulted encoding minimizes the sum of quantization error in a measure while the sum of note-lengths is equal to the number of beats. In the experiment with 16 children songs, we obtained the improved result in which measure finding was complete, the accuracy of encoding for note-length and pitch was about 91.3 and 86.7, respectively.

**Key Words** : Automatic Music Transcription, Monophonic, Time-varying Tempo, TV Filter, Pitch Smoothing, Measure Finding, Optimal Encoding of Note-length

## 1. 서 론

접수일자: 2013년 1월 21일

심사(수정)일자: 2013년 4월 3일

게재확정일자 : 2013년 4월 11일

† Corresponding author

본 논문은 한국연구재단 중견연구자지원사업(No.2012-0008726)에서 지원하여 연구하였습니다. 연구비 지원에 감사드립니다. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

최근 음성 신호 처리기술이 발전하면서 자동음악 채보 시스템을 비롯한 음악 감성 인식 및 검색 등에 대한 연구가 활발히 진행되고 있다[1-6]. 자동 음악채보 시스템은 기존의 음악에 익숙한 전문가가 직접 노래를 듣고 채보하는 방법과는 다르게, 노래가 가진 음악적 특징을 시스템이 자동으로 분석하고 인식하여 악보로 만들어 주기 때문에 작곡의 보조적인 수단으로 활용할 수 있다. 또한 자동 음악 채보 시스템의 결과는 악보를 활용한 음원 검색, 악보기반의 음악의 감성분류 및 검색, 가장 교육 프로그램 개발 등 다양하게 응용될 수 있다[7].

자동 음악채보에서 채보의 대상이 되는 음악은 한 사람

의 노래만을 대상으로 하는 단일 음(monophonic)의 경우와 여러 사람 혹은 반주가 가미된 복합 음(polyphonic)의 경우로 나누어진다. 본 논문에서는 비트 히스토그램, 스펙트랄 에너지로부터의 온셋(onset) 에너지 함수, 리듬의 통계적 모형 등을 이용하는 방법들에서 대상으로 하는 악기로 연주되는 복합음이 아니라 녹취된 단일음의 노래를 대상으로 한다[8][9].

일반적으로 사람은 악보대로 노래하지 않는다. 의도적이든 아니든 노래를 부르는 도중에 템포의 변화가 있을 수밖에 없다. 일반적으로 자동음악 채보에서는 변화하는 템포를 반영할 수도 있으며 그렇지 않을 수도 있다. 예를 들어 악보대로 노래하는 정도를 판단하여 자동으로 점수를 부여할 경우에는 변화하는 템포를 반영하여 악보와 일치 시키면 가장력을 평가하기가 곤란하다.

본 논문에서는 템포변화에 적응적인 자동채보 방법을 대상으로 한다. 이러한 템포 변화를 고려한 단일음의 채보방법은 비전문가가 흥얼거리는 허밍을 악보로 만들어 작곡하는데 도움을 줄 수도 있으며, 악보의 유사성을 이용한 내용 기반 음악검색의 전처리기로 사용될 수 있다.

기존의 템포의 변화를 반영한 자동 악보채보 방법에서는 노래를 분석하여 소절(phrase)과 마디위치를 찾아내고, 각각의 마디가 지속되는 시간, 즉 템포에 따라 음정 지속시간, 즉 IOI(Inter Onset Interval)를 정규화시켜 음길이를 부호화함으로써 가능하였다[10]. 그러나 기존의 방법에는 세 가지 문제점이 있었다. 첫째는 탐색된 마디위치의 불안정성이다. 가변 템포를 반영한 자동채보에서 마디길이 지속 시간이 템포의 변화를 의미하기 때문에 정확한 마디위치 정보는 필수적이다. 기존 방법에서는 음정, 즉 피치(pitch) 정보로부터 얻어진 IOI 만을 이용하여 마디위치를 탐색하기 때문에 마디위치 탐색의 종료시점에 따라 마디가 한 번 더 나뉘지기도 하고 덜 나뉘지기도 하는 문제가 발생하였다. 본 논문에서는 음정뿐만 아니라 음원의 에너지 정보를 함께 고려하고 박음종 등이 제안한 규칙기반 방법[6]과 융합하여 마디탐색의 안정성을 향상시켰다. 둘째로 음정 부호화에 있어서의 파편화(fragmentation)현상에 따르는 문제이다. 즉 악보 상에서 동일한 음이라도 부르는 사람의 음정이 불안한 경우에 여러 개의 짧은 IOI들로 나누어져서 음길이 부호화에 악영향을 주게 된다. 이를 최소화하기 위해서는 추출된 음정 데이터의 내재된 큰 불연속은 최대한 유지시키면서 평활화 시키는 과정이 필요하다. 이를 위해 형아영 등은 SIDE(Stabilized Inverse Diffusion Equation) 방법을 이용한 평활화를 제안하였다[5]. 그러나 이 방법은 반복횟수를 무한히 많게 하면 모두 하나의 음정으로 수렴하고 그렇지 않으면 어떻게 나누어질지 모르기 때문에 정지조건을 설정하는데 어려움이 있었다.

세 번째 문제는 모든 마디의 음표들의 합이 가설검정을 통해 얻어진 일정 박자로 구성되지 못하는 문제점이다. 템포를 반영한 IOI를 음길이로 부호화 할 경우 양자화 오차를 포함하게 되며, 한 마디 안에서 이들 양자화 오차의 합이 0이 아니기 때문에 음길이가 박자 수를 넘기도 하고 부족하기도 한다. 기존 방법은 이 문제를 박자타입에 맞게 임기응변식(heuristic)으로 오차가 가장 큰 음표들을 순차적으로 조정하여 해결을 하였지만 마디 내에서 양자화 오차의 합을 최소화 하는 최적의 방법은 되지 못하였다.

이러한 문제를 해결하기 위해 본 논문에서는 음정과 에너지, 규칙기반 방법을 융합한 마디탐색 방법으로 마디 탐

색결과와 안정성을 향상시켰으며, TV(Total Variation) 필터[11]를 이용하여 추출된 음정 데이터의 내재된 큰 불연속을 유지하면서 평탄영역의 잡음을 제거하여 음정 부호화의 파편화 현상을 개선하였다. 또한 한 마디 안에서 음표의 부호화를 양자화 오차의 합을 최소화하는 방법으로 개선하여, 향상된 채보결과를 도출하였다. 그 결과 16개의 동요 음원에 기존의 방법[10]에서는 음길이 부호의 정확도는 약 89.4%, 음정 부호화 정확도는 약 84.8%였지만, 본 논문에서 제안한 방법으로 완벽한 마디위치를 탐색하였으며 음길이 부호화의 정확도 약 91.3%, 음정 부호화의 정확도 약 86.7%의 개선된 결과를 얻을 수 있었다.

본 논문의 2장에서는 개선된 템포변화를 고려한 자동채보 방법을 순차적으로 서술하고, 3장에서는 실험 및 결과 4장에서는 결론을 제시하였다.

## 2. 개선된 자동 음악 채보방법

본 논문에서 제안한 방법은 그림 1과 같이 먼저 음원(.wav) 파일에서 음정 및 에너지 특징을 추출한다. 음정 데이터는 TV 필터로 평활화 하며, 평활화된 음정 성분은 BIC 알고리즘[12]을 통하여 동일음정 구간 즉 IOI로 나누어진다. 한편 음정 데이터에 존재하는 휴지기 정보는 소절을 찾는 데 이용되며, 분할된 IOI들은 박음종이 제안한 ‘규칙기반 마디 찾기’ 방법[6]과 ‘이진분할 마디 찾기’ 방법에 활용된다. 이들 두 방법의 한 마디 탐색결과를 에너지 데이터를 활용한 ‘이진분할 마디 찾기’방법과 결합하여 다수결(majority voting)에 따라 안정된 마디위치를 제공한다. 이렇게 찾은 마디위치를 활용하여 마디연주 시간, 즉 템포를 알아내며, 마디 연주시간에 따라 IOI들을 정규화하고 가설 검증 방법에 의해 박자를 구별하며 음길이를 부호화한다. 이 부호화 과정에서 한 마디 내에서 양자화 오차들의 합이 0이 아니기 때문에 제안된 방법으로 한 마디 안에서 양자화 오차의 합을 최소로 보정하여 음길이를 부호화하고, 상대음정 부호화 방법을 이용하여 음정을 부호화 한다.

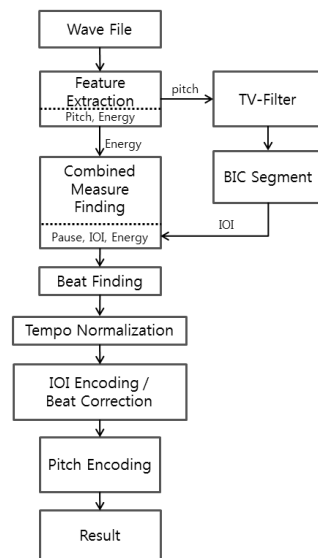


그림 1. 제안된 가변 템포 음악의 채보방법  
Fig. 1. Proposed transcription method for music with varying tempo

## 2.1 음정 추출

음악의 최소 단위인 개별 음들을 찾아내는 일은 자동 음악 채보의 성능을 좌우하는 기본적인 작업 중 하나이다. 따라서 개별 음절을 찾는 데 기초가 되는 음정 추출은 자동 채보의 정확성을 위해 중요하다. 본 논문에서는 정확한 음정 추출을 위해 Praat 프로그램을 이용하였다. 그림 2는 채보를 위해 녹음된 음원파일에서 Praat 프로그램을 이용해 음정 성분과 에너지 성분을 추출하는 모습을 보여주고 있다 [13].

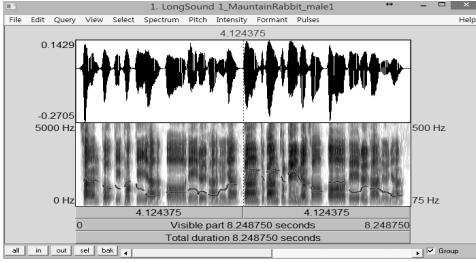


그림 2. Praat 프로그램을 이용한 음정 성분 추출  
Fig. 2. Pitch extraction using Praat program

## 2.2 TV 필터를 이용한 음정 평활화

이전 단계에서 추출된 음정 데이터로 부터 직접 음절분할(segmentation)을 시도 하게 되면 많은 불편화가 일어나게 된다. 이러한 음정의 불편화 현상은 신호에 잡음이 섞이는 경우나 노래를 부르는 사람의 음정이 불안정할 때 많이 일어난다. 기존의 방법에서는 이런 불편화 현상을 보완하기 위해 마디 안의 미리 설정된 최소 음길이 이하의 IOI들은 이전의 긴 IOI에 병합하여 음길이 부호화의 부정확성을 극복하려 하였다. 이러한 방법은 음길이에 대한 정확성을 높이는 하나의 해결책일 수 있지만 최소 음길이 설정이 임기응변적이어서 음길이 부호화도 부정확해지며, 음정을 음계로 옮기는 부분에서 강제로 병합된 IOI들 때문에 대표음정 산출에 어려움을 주어 결국 음정부호화에도 부정확성을 초래한다.

이러한 문제를 효과적으로 해결하기 위해서 본 논문에서는 TV 필터를 이용하여 음정 성분을 평활화 하였다. 식 (1)에서와 같이 잡음이 섞인 신호  $y$ , 원래 신호  $x$  잡음 신호를  $n$  이라 하자.

$$y = x + n, \quad y, x, n \in R^N \quad (1)$$

TV 필터 알고리즘은 식 2와 같은 목적함수를 최소화하는 원래신호의 추정치  $\hat{x}$ 를 탐색하는 과정을 의미하며,  $D$ 는 1차원 신호의 구배(gradient)에 해당하는 행렬이다. TV 값이란 연속되는 신호사이에 얼마나 많은 변화가 있는지 측정하는 측도로 식 (1)의 두 번째 항이다. [11]

$$\mathcal{J}(x) = \|y - x\|_2^2 + \lambda \|Dx\|_1 \quad (2)$$

본 논문에서는 식 (3)-(5)와 같이, 쌍대변수(dual variable)를 사용하며, 두 개의 변수를 갖는 함수에서 한 변수의 절대치가  $\lambda/2$ 을 넘지 못하게 제한하는 함수  $clip(\cdot, \lambda/2)$ 를

포함하는 반복적인 알고리즘을 이용하였다.

$$x^{(i+1)} = y - D^t z^{(i)} \quad (3)$$

$$z^{(i+1)} = clip(z^{(i)} + \frac{1}{\alpha} D^t x^{(i+1)}, \frac{\lambda}{2}) \quad i \geq 0 \quad (4)$$

$$z^{(0)} = 0, \quad \alpha \geq \max_e ig(DD^t) = 4 \quad (5)$$

TV 필터 알고리즘은  $\lambda$ 를 조절하여 평활화의 정도를 조절할 수 있으며, 음정 데이터의 잡음 레벨이 클수록 큰  $\lambda$ 값이 바람직하다. 본 논문의 실험에서는  $\lambda=0.1$ ,  $\alpha=4$ 로 고정하였다.

서론에서 언급한 바와 같이 형이영 등은 음정 평활화를 위해 SIDE를 적용하였다. 그러나 이 방법은 반복횟수에 따라 평활화 정도가 달라지는 단점이 있다. 즉 반복횟수를 무한히 늘리면 궁극적으로 모든 음정 데이터가 한 음계로 수렴하고 그렇지 않으면 데이터의 특성에 따라 다양한 개수의 음정으로 분할되는 성질을 가지고 있어 균일한 반복횟수를 결정하기 어려움이 있다.[14]

## 2.3 BIC를 이용한 IOI 추출

개별 음 분리를 위해서는 추출된 음정 데이터에서 음성 구간과 비음성(pause) 구간을 구분하고, 음성구간에서는 음정이 동일하게 유지되는 IOI들을 추출하는 구간분할(segmentation)이 필요하다. 본 논문에서는 이를 위해 기존의 방법에서 활용된 BIC(Baysian information criteria)알고리즘을 이용하였다[11]. 이 방법은 박은중 등의 GMM(Gaussian mixture model)에서 나타나는 시간 축상의 분할의 불일치 현상도 일어나지 않으며[6], 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

$$R(i) = N \cdot \log \Sigma - N_1 \cdot \log \Sigma_1 - N_2 \cdot \log \Sigma_2 \quad (6)$$

와 같이 정의 된다. 식 (6)에서  $N_1$ 은 첫 번째 세그먼트의 데이터 수이며,  $N_2$ 는 두 번째 세그먼트의 데이터 수이다. 또한  $\Sigma$ ,  $\Sigma_1$ ,  $\Sigma_2$ 는 각각 전체, 첫 번째 세그먼트, 두 번째 세그먼트의 공분산 행렬이다. 분할 위치  $i$ 에서 BIC는

$$BIC(i) = R(i) - \lambda P \quad (7)$$

와 같으며, 식 (7)에서  $P$ 는 패널티 항으로

$$P = \frac{1}{2} (d + \frac{1}{2} d(d+1)) \log N \quad (8)$$

과 같다. 식 (7)와 식 (8)에서  $\lambda = 1$ 이며  $d$ 는 데이터의 차원으로 피치 데이터의 경우 1차원이다. 최적의 분할 경계는 최대우도추정에 의해 식 (9)와 같이 결정할 수 있다.

$$\hat{t} = \operatorname{argmax}_i BIC(i) \quad (9)$$

만약 여러 개의 음절분할이 요구되는 경우에는 표 1과 같은 순차적인 알고리즘을 적용할 수 있다.

표 1. BIC 기반 음절 경계 탐색 알고리즘

Table 1. Syllable boundaries search algorithm based on BIC

<b>step 1</b>	Interval initialization $[a, b] : a = 1; b = 2$
<b>step 2</b>	Boundaries search using BIC in $[a, b]$
<b>step 3</b>	If there is no boundary in $[a, b]$ then $b = b + 1;$ Else $a = \hat{t} + 1; b = a + 1;$ ( $\hat{t}$ is boundary)
<b>step 4</b>	<b>step 2</b> repeat

제안된 방법에서는 분할된 음절을 개별 음으로 보고, 개별 음이 유지되는 구간을 IOI, 즉 동일음정 구간으로 간주한다.[10]

### 2.4 마디 검출 및 박자 검출방법

본 논문의 마디 검출 단계에서는 음정 추출 단계에서 찾을 수 있는 비음성 구간(pause)정보와 BIC를 이용한 IOI 정보, 음원으로부터 추출한 에너지 정보를 이용한다.

본 논문의 소절(phrase) 구분방법으로는 기존의 방법을 활용하였다[10]. 음악 전체에 소절수는 소수(prime number)의 형태로 존재하며 한 소절이 끝나면 비교적 큰 휴지기를 가진다. 따라서 2<sup>n</sup>개의 마디를 가지는 2소절 타입 음악의 경우는 노래의 1/2 위치에 긴 휴지기가 나타나며, 3<sup>n</sup>개의 마디를 가지는 3소절 타입의 음악의 경우 1/3 지점과 2/3 지점에 긴 휴지기가 나타난다. 즉 긴 휴지기들이 곡 전체상에서의 어떤 위치에 있는가에 따라 어떤 소절 타입의 노래인가를 알아낼 수 있으며, 소절을 분리해 낼 수 있다.

일단 소절의 위치를 찾아내면 한 소절에 2<sup>m</sup>개의 마디가 존재하며 이들 마디위치를 찾아낸다. 본 논문의 개선된 마디 탐색 방법은 세 가지의 방법을 융합하였다. 첫 번째는 분할된 음정성분, 즉 IOI를 이용하여 이진구간 분리 알고리즘을 적용하는 방법으로 기존의 방법 그대로이다[10]. 두 번째는 음원으로부터 추출된 소절내의 에너지 성분을 이용하여 표 2과 같이 에너지 활용 이진구간 분리 알고리즘을 적용하여 예상 마디 위치를 찾는 방법이다. 에너지를 활용한 이진 구간분리 알고리즘에서는 소절내의 에너지가 문턱치(소절의 에너지 평균치) 보다 작은 구간을 휴지기로 보고 이들 휴지기의 끝부분들을 기준으로 이진 구간 분리 알고리즘을 적용한다. 마디를 찾는 세 번째 방법은 소절내의 음표의 조합 규칙을 활용한 것으로 박은중 등이 제안한 방법이다[6]. 이 방법에서는 IOI들에 음표의 조합 규칙들을 적용하여 마디위치를 추정해내며 본 논문에서는 이 규칙들의 적용범위를 각각의 소절 내로 제한하여, 오류가 전체 노래로 전달되는 것을 방지하였다. 즉 박은중 등의 방법에서 한번 마디 탐색의 오류가 발생하면 이후에 연속적으로 오류가 생길 수 있는 가능성이 있는데, 본 논문에서는 한 소절 내에서 규칙이 적용되기 때문에 연속되는 소절에 오류가 전파되지 않는다는 의미이다.

개선된 방법에서는 이상의 세 가지 방법으로 찾아진 예상 마디 위치 중 과반수이상, 즉 두 가지 이상의 방법으로 찾은 예상 마디 위치가 동일한 위치에 있을 때 이를 최종마디라 결정하였다.

표 2. 에너지 활용 이진구간 분리 알고리즘

Table 2. Binary interval separation algorithm with energy

<b>step 1 (Initialization)</b>	Detection of end point Pause Intervals using Energy = $E$ ; Number of Measure = 1, Measure Length = Total Length of Song = $T$ ;
<b>step 2 (Recursive)</b>	Find center position of current level's Measures from set $E$ ;  If all the measures have center position near point in $E$ then binary interval separation with $E$ , repeat <b>Step 2</b> ;  Else go to <b>step3</b> ;
<b>step 3 (End)</b>	Output : Previous level's searched Measure points

마디분할 과정을 거치면 각각의 마디의 길이 즉 마디의 지속시간, 즉 템포 변화를 알 수 있다. 동일한 음표라도 템포에 따라 발생되는 기간이 달라지기 때문에 템포변화에 무관한 음길이 부호화를 위해서는 음길이를 정규화 할 필요가 있다. 본 논문에서는 기존의 방법과 같이 모든 마디의 길이의 평균값  $D_{ave}$ 를 구하고 이를 기준으로 식 (10)과 같이 IOI를 조정하였다. 조정된 IOI 들은 템포에 따라 원래 IOI를 정규화 함으로써 템포변화에 무관한 음길이 부호화가 가능하다.

$$IOI'_i = \frac{D_{ave} \cdot IOI_i}{D} \tag{10}$$

본 논문의 박자 검출에 있어서도 기존의 가설검증 방법을 적용하였다[10]. 가설검증을 이용한 박자 검출 방법에서는 4박자 노래라는 가설을 바탕으로  $D_{ave}$ 를 4로 나누어 4분 음표의 길이를 가정하고 이를 바탕으로 모든 음표의 지속시간을 산출한다. 이렇게 산출된 음길이를 활용하여 조정된 IOI, 즉 IOI'들을 임시 부호화하여 각각의 마디를 4박자의 가정에 적합하지, 3박자 가정에 적합한지를 결정한다. 이러한 각 마디의 결정 사항은 전체 노래의 마디에 대해 누적하고 다수결에 의해 4박자 또는 3박자 노래임을 결정한다.

또한 2/4박자와 4/4박자의 구별은 4/4박자를 가정한 임시 부호화 과정에서 노래 전체에서 16분 음표가 존재하지 않으면 2/4박자로 간주하였다. 이는 만약 16분 음표가 임시 부호화 과정에서 나타났다면 2/4박자에서는 32분 음표에 해당되어 최소 음이 16분 음표보다 작아지게 되기 때문이다.

이러한 방법은 전체 노래의 마디에 대해 각 마디의 결정 사항을 누적하고 있기 때문에 박자의 변화가 없다는 가정을 바탕으로 하고 있다. 또한 16분 음표보다 짧은 음표는 존재하지 않는다는 가정을 바탕으로 하고 있다.

### 2.5 음길이의 부호화 및 보정

본 논문의 방법에서 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  $0.25(\text{♩})$ ,  $0.5(\text{♩})$ ,  $0.75(\text{♩})$ ,  $1(\text{♩})$ ,  $1.5(\text{♩})$ ,  $2(\text{♩})$ ,  $3(\text{♩})$ ,  $4(\text{♩})$ 를 곱하면 음길이 부호화에 사용될 표준 음길이가 산출되며, 조정된  $IOI_i'$  또는 휴지기들의 길이를 이들 표준길이에 비교하는 방식으로 음표부호화가 진행된다[10].

음길이 부호화는 마디 내에 부호화된 음길이의 합이 전체 박자수와 일치하고 양자화 오차의 합이 최소가 되도록 하는, 제약조건이 있는 식 (11)와 같은 최적화 문제로 볼 수 있다.

$$\min \sum_{i=1}^n |IOI_i' - NT_i| \quad s.t. \quad (11)$$

$$\sum_{i=1}^n NT_i = \text{Number of beats}$$

식 (11)에서  $NT_i$ 는 마디 내에  $n$ 개의  $IOI_i'$ 에 대한 음길이 부호화의 결과의 표준길이, 즉 양자화된 길이이며 이들의 합은 박자수와 일치해야 한다.

그러나 마디 내의  $IOI_i'$ 들을 개별적으로 가장 가까운 표준 음길이에 부호화하는 방식은 식 (11)의 양자화 오차의 합을 최소로 하지 않을 수 있다. 또한 식 (11)의 제약조건을 만족시키지 못하게 되어, 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.

따라서 기존 방법의 음표보정 과정에서는 부호화된 음표의 합이 박자 수에 모자를 경우에는 음표의 길이를 늘려야 하는데 오차가 가장 큰, 즉 부호화된 표준음 보다 실제길이가 가장 긴  $IOI_i'$ 의 음표를 한 단계 늘려주었다. 반면에 그 반대의 경우에는 음의 오차 값을 가지는  $IOI_i'$  중에 가장 작은 오차를 가지는 (절대치로 볼 때는 가장 큰 오차)  $IOI_i'$ 를 찾아 한 단계 줄여주며, 이러한 보정에 의해 경우에 따라서는 박자수를 넘기도 하고 부족하기도 하기 때문에 음표길이의 합이 박자수와 같을 때 까지 이 과정을 반복하였다. 따라서 임기응변적인 기존의 방법은 식 (11)의 제약조건은 만족하지만 양자화 오차의 합을 최소화시킨다고 말할 수 없다.

표 3의 제안된 방법에서는 일단 기존방법과 같이 개별  $IOI_i'$ 들의 음길이 부호화를 진행하고, 이들 임시 부호화된 음길이가 4분 음표이하인 경우, 현재 개별 부호화된 음 길이와 이들 길이를 한 단계 줄이거나 늘린 음표를 포함하는 가능한 탐색음표 조합을 구성하였다. 예를 들어 한 마디에 4분 음표이하로 개별 부호화된  $IOI_i'$ 이 4개라면  $3^4$ 개의 음표조합을 탐색대상으로 하여 식 (4)의 제약조건을 만족하며 최소의 오차 합을 만족하는 음표 열을 찾아냈다. 여기서 4분음표 보다 긴 음표는 고려하지 않은 이유는 점 4분 음표 이상의 경우 음표를 한 단계 길게 하거나 짧게 하여 탐색조합을 구성할 경우 변화량이 커 불필요한 조합을 탐색하기 때문이다.

제안된 방법은 개별  $IOI_i'$ 들을 일단 가장 가까운 음표로 부호화하여 최적해 근방에 접근시키며 이 최적해 근방에서 외란을 통해 탐색집합을 구성하고 이들 중에 제약조건을

만족하는 최적 해를 찾아가는 수치적 최적화 방법이다.

표 3. 음표 조합 구성 및 음길이 보정 알고리즘  
Table 3. Note pool construction and note length correction

step 1	All $IOI_i'$ s in Measure encode $IOI_i'$ into Nearest standard Note length
step 2	All of step 1's result(Nearest standard Note length) If under Quarter Note length low or upper note length combination to construct note pool else take current note length combination to construct note pool
step 3	Select note combination which satisfy equation (11) and the lowest sum of error in note combination from note pool
step 4	Output : final note combination

## 2.6 음정의 부호화 방법

음정 부호화 방법에는 절대음정을 이용하는 방법과 상대 음정을 이용하는 방법이 있다[15]. 절대음정을 이용할 경우에는 노래의 건(key)을 고려하여 부호화할 수 있으며, 상대 음정일 경우에는 건 정보는 무시되고 모든 음이 다장조를 기준으로 부호화 된다. 본 논문에서는 형아영 등이 제안한 음정부호화 방법을 활용하였다[5]. 이들은 사람이 음정을 인지할 때와 같이 외부에서 주어지는 기준 음에 의존하여 음정을 판별하는 상대음정 부호화 방법을 사용하였다. 즉 그림 4와 같이 음정 데이터와 적용할 21개 모델의 상대 음과의 유사성을 측정하여 가장 높은 유사도를 음정모델을 찾고 이를 이용하여 그림 4과 같이 12음계에 대한 음정부호화를 시도하였다.

각  $IOI_i'$ 의 대표음정을 결정하는 과정에서 노래하는 사람의 음정불안의 영향을 최소화하기 위해 외톨이(outlier) 데이터를 배제한 알파 절삭 평균(trimmed mean)을 이용하였다. 실험에서 알파 값은 20%로 하였다.

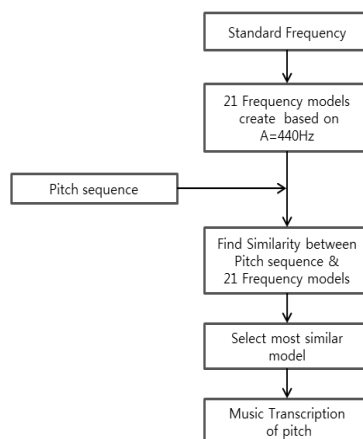


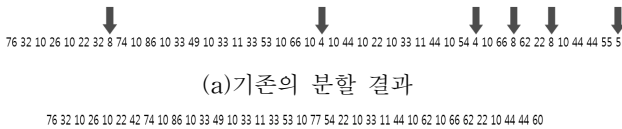
그림 4. 상대 음정을 이용한 음정부호화  
Fig. 4. Relative pitch encoding

### 3. 실험 및 결과

본 논문의 실험에서 사용된 음원은 14곡의 동요음원과 2곡의 애국가 음원으로 총 16개의 음원이었으며, 악기소리 없이 오직 여성과 남성의 목소리만 녹음되었다.

본 실험에서 채보된 결과를 원본 악보와 비교했을 경우 표 3과 같이 전체적으로 음길이의 경우, 약 91.1%의 정확도를 음정은 약 86.7%의 정확도를 보였다. 이는 기존의 방법 [10]에 비해 음길이 부호의 정확도는 약 1.7%, 음정 부호화 정확도는 약 1.9% 향상된 결과이다. 기존의 방법의 실험에서와 같이 음길이의 부호화 결과에서, 같은 음정이 반복되어 하나로 합쳐지는 경우는 전체 음길이의 합이 맞을 경우 옳은 경우라고 간주하였다. 음정 부호화의 오류가 발생한 경우도 모두 반응차이 정도만 존재하는 것을 확인하였다.

본 논문에서 제안된 방법으로 음정 파편화 문제는 그림 5와 같이 개선됨을 확인하였다. 그림 5는 음정 성분을 BIC 알고리즘을 통해 IOI로 분할된 길이 정보이다. 그림 4의 (a)를 보면 화살표로 표시된 IOI들은 작게 파편화 보정이 필요하게 된 부분이었지만, 그림 5의 (b)를 보면 TV 필터를 거쳐 작게 파편화된 IOI들이 모두 제거되는 것을 확인할 수 있다. 따라서 제안된 방법의 개선사항이 음길이 및 음정 부호화의 정확도에서 향상시킨 것으로 나타났다고 판단된다.



(a) 기존의 분할 결과

76 32 10 26 10 22 42 74 10 86 10 33 49 10 33 11 33 53 10 77 54 22 10 33 11 44 10 62 10 66 62 22 10 44 44 60

(b) TV-Filter 적용한 분할 결과

그림 5. TV-Filter를 적용한 결과와 이전 결과의 비교

Fig. 5. Compare with None Filtered and Tv-Filtered result

그림 6는 음원 song12(학교중)의 채보 결과로 제안된 방법의 성능개선 효과를 찾을 수 있다. 그림 6의 (a)부분의 2소절 첫째 마디의 음길이가 원 악보와 다른데 반해 본 논문에서 제안한 방법을 적용한 그림 6의 (b)의 2소절 첫째 마디에는 개선된 음길이 부호화 결과를 보여주고 있다.



(a) 기존의 방법

(b) 제안한 방법



(c) 원곡 악보

그림 6. 채보된 결과 비교

Fig. 6. Note transcription result

표 4은 제안된 방법으로 채보된 전체결과를 보여주고 있다. 그림에서 알 수 있듯이 기존 방법에서 전체 노래에서 2개의 음악의 마디구조를 파악하지 못했던 것에 비해, 개선된 방법에서는 모든 노래의 마디구조를 정확하게 파악함을 확인할 수 있다. 이는 세 가지 방법으로 결정된 마디위치를 다수결에 의해 융합하여 마디 탐색의 성능이 개선되었음을 의미한다.

표 4. 음길이와 음정의 정확도

Table 4. Note length and note scale accuracy

Songs	Measure detection methods		Note of length(%)	Note of code(%)
	Previous	Proposed		
song 1	correct	correct	94	90
song 2	-	correct	96	92
song 3	correct	correct	93	90
song 4	correct	correct	91	86
song 5	correct	correct	92	82
song 6	correct	correct	89	85
song 7	correct	correct	92	84
song 8	correct	correct	90	87
song 9	correct	correct	89	94
song 10	correct	correct	88	84
song 11	correct	correct	89	80
song 12	correct	correct	94	81
song 13	correct	correct	90	90
song 14	correct	correct	93	87
song 15	correct	correct	90	88
song 16	-	correct	89	86
Total			91.33	86.67

### 4. 결론

사람이 부르는 노래와 같이 단일음원으로 부터의 변화하는 템포를 반영한 자동 채보는 허밍을 사용한 음원검색, 작곡 도우미 시스템 등 다양한 분야에 활용될 수 있다.

본 논문에서는 채보의 정확성을 증가시키기 위한 기존의 방법을 개선하는 세 가지 방안을 제시하였다. 첫째, TV 필터를 활용한 음정 데이터의 평활화로 BIC 알고리즘에 의한 음정분할 결과의 파편화 현상이 줄어들어, 기존 방법과 달리 인위적인 음길이 병합을 행할 필요성이 없어졌다.

또한 음정과 에너지, 규칙기반 방법을 융합한 마디탐색 방법으로 마디 탐색결과의 안정성을 향상시켰다. 뿐만 아니라 마디 내에서 음표 조합을 구성하고 이 조합내에서 탐색을 통해 음표의 합이 박자수와 같으면서 양자화 오차의 합을 최소화하는 방법으로 임시 음길이 부호화 결과를 보정하였다.

그 결과 16개의 동요 음원에서는 완벽한 마디위치를 탐색하였으며, 음길이 부호의 정확도 약 91.3%, 음정 부호화 정확도는 약 86.7%의 개선된 결과를 얻을 수 있었다.

향후 국악의 자동 채보와 동요 이외의 다른 장르의 복합 음원의 채보에 대해서도 지속적인 연구를 수행할 예정이다.

## References

- [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] Martin, "A Blackboard System for Automatic Transcription of Simple Polyphonic Music," *MIT Media Laboratory Perceptual Computing Section Technical Report No.385*, 1996.
- [3] Marolt, "A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music," *IEEE Transactions on Multimedia*, Vol. 6, No.3, pp.439-449, 2004.
- [4] Jung-yeong Jang, Design and Implementation of Song Recognition System using Fuzzy Integral, *KAIST, Master Thesis*, 1996.
- [5] Ayoung Heung, Joonwhoan Lee, "Automatic Music Transcription System Using SIDE and Genetic Algorithm," *Journal of Korea Information Processing Society*, Vol.16, No.2, 2009.
- [6] Eungjong Park, Songyi Sin, Joonwhoan Lee, "Finding Measure Position Using Combination Rules of Musical Notes in Monophonic Song," *Journal of The Korea Contents Association*, Vol.9, No.10, pp.1-12, 2009.
- [7] Jeonggyu Ji, "Melodic query processor for efficient retrieval of an audio database," *Soongsil University, PhD Thesis*, 1998.
- [8] Geoffroy Peeters, "Time Variable Tempo Detction and Beat Marking," *International Computer Music Association*, Barcelona, Spain, 2005.
- [9] 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.
- [10] Youngho Ju, Baniya Babukaji, Joonwhoan Lee, "Automatic Music Transcription Considering Time-Varying Tempo," *Journal of The Korea Contents Association*, Vol.12, No.11, pp.9-19, 2012.
- [11] Ivan W. Selesnick and Ilker Bayram, "Total Variation Filtering", Available: [http://eeweb.poly.edu/iselesni/lecture\\_notes/TV\\_filtering.pdf](http://eeweb.poly.edu/iselesni/lecture_notes/TV_filtering.pdf), February, 2010, [Accessed: January 14, 2013].
- [12] 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.
- [13] Praat, Available: <http://www.praat.org>, [Accessed: January 14, 2013].
- [14] Ilua Pollak, Alan S. Willsky and Hamid Krim, "Image Segmentation and Edge Enhancement with Stabilized Inverse Diffusion Equations", *IEEE Transactions on Image Processing*, Vol. 9, No.2, pp.256-266, 2000.
- [15] Korean Society for Music Perception and Cognition, *Perseption and Congnition of Music*, Eumaksekye, 2005.

## 저 자 소 개



### 주 영 호 (Youngho Ju)

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

관심 분야 : 음성처리, 멀티미디어  
Phone : +82-10-2308-8312  
E-mail : nightmute@jbnu.ac.kr



### 이 준 환 (Joonwhoan Lee)

- 1980년 : 한양대학교 전자공학과 (공학사)
- 1982년 : 한국과학기술원 전자공학과 (공학석사)
- 1990년 : 미국 미주리대학 전기 및 컴퓨터공학과(공학박사)
- 1990년 ~ 현재 : 전북대학교 전자정보공학부 교수

관심 분야 : 영상처리, 컴퓨터 비전, 인공지능  
Phone : +82-10-9855-2406  
E-mail : chlee@jbnu.ac.kr