

단일 음원 노래에서 음표의 조합 규칙을 이용한 마디 위치 찾기

Finding Measure Position Using Combination Rules of Musical Notes in Monophonic Song

박은종*, 신송이**, 이준환***

한국전자통신 연구원*, 전북대학교 컴퓨터공학과**, 전북대학교 전자정보공학부***

En-Jong Park(for0511@etri.re.kr)*, Song-Yi Shin(blueseraph@empal.com)**,
Joon-Whoan Lee(chlee@chonbuk.ac.kr)***

요약

본 논문은 한 마디 내에서 나타나는 음표들을 조합했을 때 규칙적인 배수관계가 있음을 이용하여 단음으로 구성된 노래에서 정확하게 마디 위치를 찾는 방법을 제안한다. 제안된 방법에서는 먼저 노래의 개별음의 길이들을 찾아내고, 이들 개별음 길이들의 배수관계를 적용한 규칙들을 이용한다. 이렇게 찾아진 마디정보는 자동 음악 채보 시스템의 사전 정보로 활용될 수 있는 박자 및 빠르기 등을 위한 원천적인 정보를 추출하는데 활용될 수 있다. 남성 및 여성의 단음으로 구성된 12곡의 노래들에 대한 실험 결과 1곡을 제외한 11곡 노래들의 마디위치를 정확히 검출 했으며, 이렇게 찾은 마디 정보와 음악적 이론을 이용하여 곡의 박자 및 빠르기 정보 등을 추출할 수 있다.

■ 중심어 : | 음표의 조합 규칙 | 자동 음악 채보 | 마디 찾기 | 박자 및 빠르기 검출 |

Abstract

There exist some regular multiple relations in the intervals of notes when they are combined within one measure. This paper presents a method to find the exact measure positions in monophonic song based on those relations. In the proposed method the individual intervals are segmented at first and the rules that state the multiple relations are used to find the measure position. The measures can be applied as the foundational information for extracting beat and tempo of a song which can be used as background knowledge of automatic music transcription system. The proposed method exactly detected the measure positions of 11 songs out of 12 songs except one song which consist of monophonic voice song of the men and women. Also one can extract the information of beat and tempo of a song using the information about extracted measure positions with music theory.

■ keyword : | Combination Rules of Musical Notes, Automatic Music Transcription | Finding Measure Position | Extraction Beat and Tempo of Song |

I. 서 론

최근 음성 신호 처리 분야가 발전하면서 자동 음악

채보 시스템을 비롯한 음악 감성 인식 및 검색 등에 대한 연구가 활발히 진행되고 있다[1]. 자동 음악 채보 시스템은 기존의 음악에 익숙한 전문가가 직접 노래를 듣

고 채보하는 방법에 비하여 시창자의 노래가 가진 음악적 특징을 시스템이 자동으로 인식하여 일반인도 쉽게 노래를 악보화 할 수 있는 보조적인 수단으로 활용되고 있다. 또한 자동 음악 채보 시스템의 결과물을 이용하여 내용기반 음원 검색을 비롯한 음악의 감성 분류 등 다양한 응용 분야로 확장되고 있다[2][3]. 실제로 2008년 9월에 출시된 Apple의 iTunes 8은 저장된 음악 라이브러리에서 play list를 자동 생성할 수 있는 'Genius' 기능을 탑재 하였다. 이는 라이브러리에서 선택된 음원과 유사한 종류의 음원을 모아 새로운 play list를 자동으로 생성해 주는 기능으로 일종의 내용기반 음원 검색에 해당된다. 또한 2008년 9월 출시된 Microsoft 사의 Zune 신 모델의 경우는 라디오에서 나오는 음악을 인식해 곡을 설명해주고 다운로드를 받을 수 있는 기능과 특정음악가를 설정하면 비슷한 성향을 갖는 다른 음악을 추천 받을 수 있는 서비스를 지원한다. 이러한 Zune 신 모델의 새로운 기능은 음악을 인식하는 기술과 음악 분류 기술을 필요로 한다. 또한 일본에서는 음악을 색상이나 키워드로 검색, 구입할 수 있는 사이트가 있으며 사용자는 무료로 인디 음악을 감상하고, 감상한 느낌을 색상이나 키워드로 곡에 연관 지을 수 있다. 이후 또 다른 청취자가 색이나 키워드를 선택하는 것으로, 관련되어진 곡을 검색해 감상이나 구입, 음악을 보다 직접적인 감성으로 검색할 수 있고, 자신의 이미지에 맞는 새로운 곡을 선택할 가능성을 제공하고 있다[4].

이러한 응용 분야의 핵심적인 근본 기술은 자동 음악 채보 시스템이라고 말할 수 있으며, 단일 음원에 대한 대표적인 자동 음악 채보 시스템은 creativedectors[5], recognisoft[6] 등으로 말할 수 있다. 이들은 사람이 부른 단일 음을 분석하여 다양한 응용프로그램인 PitchScope, NoteDetection 등을 제공한다. 단일 음원에 대한 채보 시스템을 확장한 복합 음원에 대한 대표적인 자동 음악 채보 시스템은 [7-9]등으로 이 시스템들은 Blackboard 구조를 적용하여 피아노의 복합 선율이 포함된 연주곡을 비롯한 다양한 악기로 구성된 복합 음원을 분석, midi파일 형태로 결과물을 생성시켜 주는 좀더 복잡한 곡에 대한 변환이 가능하다.

그러나 이러한 자동 음악 채보 시스템들은 원곡의 발

성 및 연주와 가장 흡사한 결과물 생성만을 목적으로 하기 때문에 만족할만한 결과물을 얻기 위해 원곡의 박자 및 빠르기에 대한 사전 정보의 입력을 필요로 한다. 즉, 곡의 박자에 대한 정보를 알아내기 위해서는 박자의 최소 구별 단위인 마디 위치를 찾아내야 하는 복잡한 과정과 계산 량이 필요하기 때문에 대부분의 자동 음악 채보 시스템들은 박자 및 빠르기 등에 대해 파라미터 형식의 사전 정보를 요구하고 있다.

본 논문에서는 자동 음악 채보 시스템의 이러한 사전 정보를 자동으로 추출하기 위한 방법을 제안한다. 사람이 발성한 단일 음원 곡들에 대해 마디의 구성 요소인 음표들의 조합규칙을 제안하고, 이를 이용하여 마디 위치를 찾는 방법과 3박자 계통 및 4박자 계통의 박자 정보 유추와 찾아진 마디 정보를 이용한 곡의 빠르기 정보 추출 방법을 제안한다.

II. 개별 음표 찾기

곡의 최소 단위인 개별 음들을 찾아내는 일은 자동 음악 채보에 있어서 성능을 좌우하는 기본적인 작업 중 하나라고 말할 수 있다.

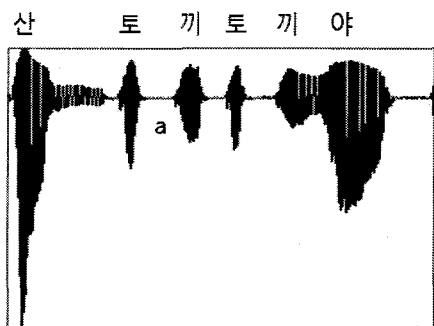


그림 1. 음성신호, 동요 "산토끼"

[그림 1]은 동요 "산토끼" 중 처음 두 마디에 해당하는 "산토끼 토끼야"에 대한 음성 신호를 나타낸 것으로 [그림 1]의 a는 사람이 노래할 때 발생하는 음성의 단절 부분인 휴지기(interphase)를 나타내고 있다. [그림 1]의 음성 신호를 자세히 살펴보면 "산"에 대한 신호와

마지막 “끼야”에 대한 신호의 파형이 하나의 발성에 대한 파형으로 보기에 모호한 부분이 존재함을 알 수 있다. 즉 “끼야”에 해당하는 파형을 하나의 개별 음으로부터 얻어진 파형인지 아닌지에 대한 결정 과정이 반드시 존재해야 한다는 것이다. 만약 “끼야”에 대한 파형을 두 개의 개별 음이 발생된 것이 아닌 하나의 개별 음이 발생된 것으로 판단한다면 자동 음악 채보 시스템의 성능은 크게 떨어지게 될 것은 자명할 것이다.

본 장에서는 이러한 개별 음을 찾기 위한 방법으로 피치 샘플들의 표준편차 방법과 GMM(Gaussian Mixture Model)[10]을 이용한 방법을 제안한다.

1. 피치 샘플들의 표준편차

개별 음을 찾기 위한 첫 번째 과정은 휴지기를 개별 음들 사이의 구분 지점으로 보고 휴지기가까지 포함하는 개별 음들의 후보들을 찾아내는 것이다[11]. [그림 2]는 [그림 1]의 음성 파형에 대해 휴지기를 이용하여 휴지기를 포함한 개별 음들의 1차 구분을 나타낸 것이다.

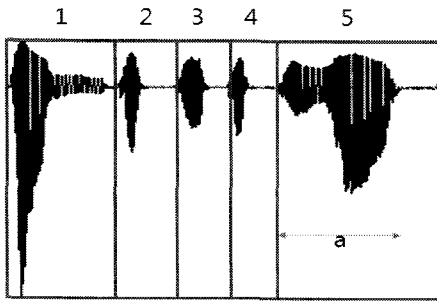


그림 2. 휴지기를 포함한 개별 음들의 구분 지점

[그림 2]의 결과에서도 알 수 있듯이 개별 음들은 노래를 부르는 사람의 발성 형태에 따라 두 개 이상의 음들이 섞여서 발생될 수도 있고([그림 2]의 5부분), 오직 하나의 음만이 발생될 수도 있다. 이러한 개별 음들의 혼합 발생 유무는 휴지기를 이용하여 나누어진 개별 음의 피치 샘플 값들에 대한 표준 편차를 통해 알 수 있다. 즉, 사람의 발성이 정확히 하나의 음만을 발생하였다면 피치 샘플의 표준 편차는 크지 않을 것이고 만약 복수개의 음들이 발생되었다면 그 속에서는 복수개의

피치샘플 분포로 말미암아 표준 편차가 크게 나타난다고 가정한다.

본 논문에서는 1차 구분한 개별 음들의 피치를 0.01초 단위로 샘플링 하여 표준편차를 구하였다.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - m)^2}{n}} \quad (1)$$

수식 (1)은 피치 샘플들의 표준 편차를 나타낸다. 수식(1)의 n은 [그림 2]의 각 개별 음들의 길이 가운데 휴지기를 제외한 부분에서 피치를 0.01초 단위로 샘플링한 표본의 크기로 예를 들어 [그림 2]의 경우 5부분은 길이 a 부분의 피치샘플의 크기에 해당한다.

[표 1]은 [그림 2]의 개별 음표들의 샘플링 된 피치 값의 평균과 표준 편차를 나타낸 것으로 [그림 2]의 5부분의 결과에서 피치샘플 표준 편차가 현저하게 큰 것으로 나타났다. 이는 앞에서 우리가 가정한 하나의 파형에 다수의 음이 혼재되어 나타났다는 것을 의미한다.

표 1. 개별 음들의 평균 및 표준편차

| | 1 | 2 | 3 | 4 | 5 |
|----------|--------|--------|--------|--------|--------|
| m | 200.72 | 160.28 | 170.41 | 207.96 | 146.51 |
| σ | 2.82 | 5.46 | 2.45 | 4.78 | 17.38 |

본 논문에서는 [표 1]의 1, 2, 3, 4번의 개별 음들과 같이 표준 편차의 범위가 T보다 작으면 하나의 음이 안정적으로 발생된 것으로 간주 하고, [그림 4]의 Part 1에서와 같이 해당 음을 개별 음표 그룹 스택에 저장하도록 한다.

2. 피치 샘플들의 이동 표준편차와 GMM

[표 1]의 5번음과 같이 표준 편차의 범위가 T보다 큰 개별 음들은 Π -1절을 통해 다수의 음이 혼재되어 있음을 판단 할 수 있으므로 이를 다시 의미 있는 독립적 개별 음들로 재분리 하는 과정이 필요하다.

재분리를 위한 첫 번째 방법은 피치 샘플들의 이동 표준편차를 이용한 방법이다.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - m_i)^2}{i}}, \quad i \ll n \quad (2)$$

수식 (2)는 혼합 음이 내재되어 있는 개별 음을 분리하기 위해 개별 음의 샘플링 된 모든 데이터들을 하나씩 누적해 가면서 이동 표준 편차를 구하는 것을 의미한다. 만약 임의의 지점에서의 표준편차가 크게 나타날 경우에는 새로운 개별 음의 영역으로 판단하여 개별 음의 분할 개수를 증가시키고, 이 지점을 분할 지점으로 가동록한다. 이렇게 가동록된 음의 분할 개수는 GMM을 이용한 정교한 분할 방법을 위한 사전 정보로 활용된다.

GMM은 주어진 데이터를 임의의 개수의 가우시안 분포의 혼합으로 모델링하는 방법이다[10]. 다음 수식은 K개의 가우시안 분포를 가지는 GMM을 나타낸 것이다.

$$P(X_t) = \sum_{j=1}^K w_{j,t} * \eta(X_t, \mu_{j,t}, \Sigma_{j,t}) \quad (3)$$

수식 (3)에서 $w_{j,t}$ 는 시간 t 에서 j 번째 가우시안 모델의 가중 계수를 의미하며, $\mu_{j,t}$ 와 $\Sigma_{j,t}$ 는 시간 t 에서 j 번째 가우시안의 평균과 분산행렬이다. η 는 가우시안 확률 분포 함수로 수식 (4)와 같이 표현된다.

$$\eta(X, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)} \quad (4)$$

수식 (3)의 K는 주어진 데이터의 가우시안 분포의 수를 의미하므로 사전에 파라미터 형식으로 결정을 해야만 한다. 본 논문에서 처리하는 데이터는 1차원의 샘플링 된 피치 값이며, 앞서 언급한 피치 샘플들의 이동 표준 편차 방법을 통해 개별 음 내에 몇 개의 음들이 혼재되어 있는지를 파악하였으므로 찾아진 수만큼의 가우시안 분포가 존재한다고 가정하여 K값을 설정하였다.

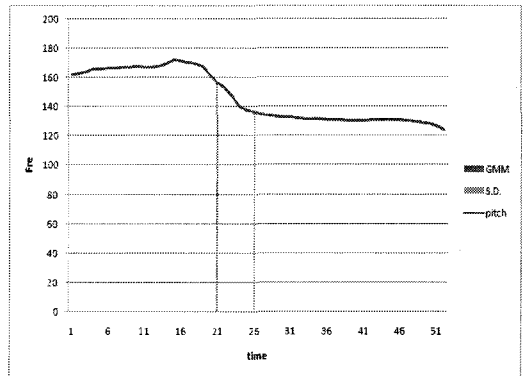
$$\begin{aligned} w_{k,t} &= (1-\alpha)w_{k,t-1} + \alpha(M_{k,t}) \\ \mu_{k,t} &= (1-\rho)\mu_{k,t-1} + \rho X_t \\ \sigma_{k,t}^2 &= (1-\rho)\sigma_{k,t-1}^2 + \rho(X_t - \mu_{k,t})^2 (X_t - \mu_{k,t}) \end{aligned}$$

where

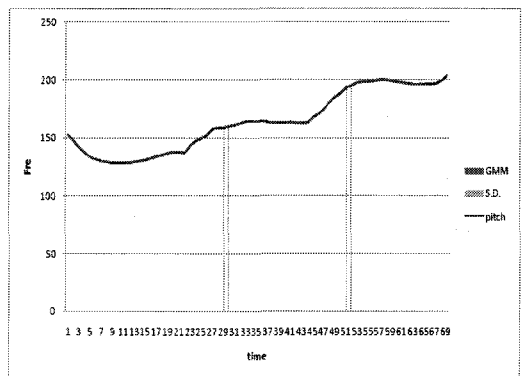
$$M_{k,t} = \begin{cases} 1: & \text{for the model which matched} \\ 0: & \text{for the remaining model} \end{cases} \quad (5)$$

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k)$$

수식 (5)는 GMM 각 계수들의 갱신을 나타낸다. 수식 (5)에서 가중치 $w_{k,t}$ 의 갱신에 있어 이전 가중치의 영향력을 반영하는 α 값은 실험을 통하여 0.4로 설정하였다. 개별음의 경계 구분은 수식 (3)의 GMM의 인접한 모드사이에서 확률밀도 값이 최소가 되는 지점을 택하여 활용하였다.



(a) 발생된 소리 안에 두 개의 개별 음들이 포함된 경우



(b) 발생된 소리 안에 세 개의 개별 음들이 포함된 경우

그림 3. GMM과 이동 표준편차를 이용한 최적 분할지점 찾기

[그림 3]은 GMM과 이동 표준편차 방법을 이용한 분할 결과들을 나타낸다. [그림 3]의 (a)는 [그림 2]의 5부분에 해당하는 개별 음의 분할 결과로 단순히 이동 표준 편차만을 이용하여 분할했을 때는 두 음의 분할 지점이 극단적인 위치에서 분할되고 있음을 알 수 있다. 이는 연속적인 발생이 이루어졌을 때 두 음의 연결 지

점인 중간부분이 아니라, 극단적으로 앞의 음이나 뒤의 음으로 치부해 버리는 단점을 내포하고 있다고 볼 수 있다. 반면에 GMM을 이용한 분할은 두 영역의 경우 연속적으로 변하는 피치의 중간지점을 찾음으로써 좀 더 정확한 분할 지점을 찾고 있음을 확인 할 수 있다. [그림 4]는 개별 음을 찾기 위한 지금까지의 과정을 보여준다.

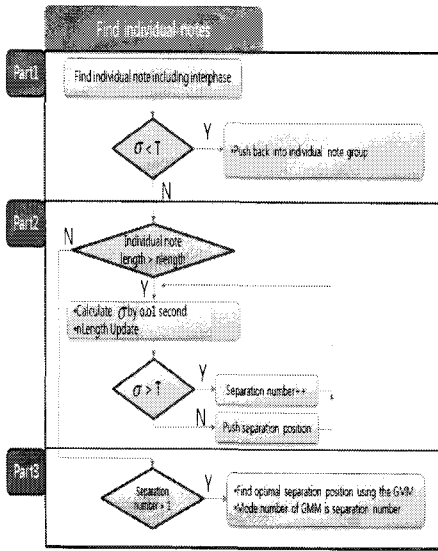


그림 4. 개별 음표 찾기

III. 음표의 조합 규칙

만약 우리가 사전에 곡의 박자와 빠르기 정보를 알고 있다고 가정한다면, 마디를 구성하는 음표들을 정확히 구분해서 분리할 수 있을 것이다. 그러나 곡에 대한 박자 정보를 알기 위해서는 마디의 분석이 우선적이다. 왜냐하면 음악적으로 하나의 마디는 곡의 박자에 대한 정보, 즉 음표들을 가지고 있기 때문이다. 이러한 이유로 곡의 박자 정보를 얻기 위해서는 마디를 정확히 찾아내는 과정이 필요하다고 말할 수 있다.

음악에서 마디는 마디를 구성하고 있는 음표들의 집합으로 말할 수 있다. 본 장에서는 마디를 구성하는 최소단위인 음표들을 조합했을 때 나타나는 규칙적인 배

수 관계를 이용하여 마디를 찾는 일련의 규칙들을 설명한다.

곡의 처음부터 시작하여 II절의 방법으로 구분된 i 번째 개별음의 길이를 T_i ($i = 0, 1, \dots, n - 1$)라 하자. 또한 이들을 i 번째 개별 음까지 누적한 음의 길이를

$$S_i = \sum_{k=0}^i T_k$$

라 하자. 만약 i 보다 큰 j 에서 $S_j/S_i = l_{ij}$ 를 만족하는 2이상의 정수 l_{ij} 가 존재한다면 누적하여 조합된 음표길이의 배수관계가 존재한다고 말할 수 있다.

1. 규칙1 - 징검다리 배수가 존재할 경우

만약 음표 조합의 길이에 징검다리 배수가 존재하면 배수관계의 마지막 종착부분이 마디의 위치이다.



(a) 실제 악보



(b) 이상적인 음표 조합 과정

그림 5. 규칙 1에 대한 이상적인 음표조합 과정

[그림 5]는 규칙 1이 적용되는 이상적인 경우의 예를 보여준다. [그림 5]의 (b)의 첫 번째 경우는 처음 등장한 8분 음표로 누적된 음길이의 배수관계를 찾는 시도를 한 경우로 $S_1/S_0 = 2$ 로 배수관계가 있음을 알 수 있다. 또한 두 번째 경우는 처음 연속적으로 발생된 8분 음표 2개의 길이, 즉 4분 음표 길이를 이용하여 배수관계를 확인한 경우로 $S_3/S_1 = 2$ 로 배수관계가 있음을 알 수 있다. 세 번째 경우는 처음 세 개의 연속된 8분음들을 누적한 점 4분 음표로 연속되는 길이의 배수관계

를 확인하는 경우로 $S_5/S_2 = 2$ 의 배수 관계를 확인할 수 있다. 또한 네 번째 경우는 처음 발생된 4개의 8분 음표길이를 누적하여, 즉 2분 음표 길이로 배수관계를 확인하는 경우로 $S_7/S_3 = 2$ 의 배수관계를 가지고 있음을 알 수 있다. 규칙 1에서 징검다리 배수란 $S_j/S_i = l_{ij}$ 뿐만 아니라 $S_k/S_j = l_{jk}$ 의 배수관계도 연속된 i, j, k 에서 존재하여 $S_k/S_i = l_{ij}l_{jk} = l_{ik}$ 의 배수관계도 성립하는 경우로 징검다리 위치는 S_j 이다. [그림 5]에서 8분, 4분, 2분 음표로 연속적으로 이어진 배수의 배수관계는 징검다리 배수의 예이며, 마디의 위치는 징검다리 배수의 마지막 종착부분 S_k 의 위치, 즉 예에서 2분 음표까지의 길이에 해당한다. 점 4분 음표로 시도하는 경우는 만약 3분 음표로 시도할 경우에도 배수관계가 존재하면 징검다리 배수관계로 볼 수 있으나 [그림 5]의 경우에는 해당되지 않는다.

하지만 사람들이 일반적으로 악보에 따라 정확하게 노래하지 않기 때문에 배수관계는 오차를 포함할 수밖에 없다. [그림 6]은 [그림 5]의 실제 경우를 도시한 것으로 가로(w)축의 숫자는 연속적으로 누적되는 개별 발생음의 길이를 의미하며 세로(h)축의 숫자는 연속적으로 나타나는 개별음의 발생 길이를 의미한다. 가로 세로 축의 숫자는 0.01초 단위로 샘플링한 경우의 샘플 수로 길이를 표현한 것이다.

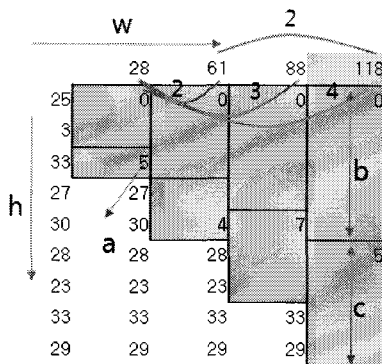


그림 6. 그림 5 노래의 실제 발생에 대한 조합 과정

즉 [그림 6]에서 개별 음들은 0.25s, 0.03s, 0.33s 등의 순서로 발생되고 있으며, 이들을 누적인 길이는 0.28s,

0.61s 등이다. 규칙 1의 적용과정에서 10(0.1ms)이하의 개별 음은 의도적으로 발생된 음으로 볼 수 없기 때문에 선행음에 연결된 음이라 가정하였으며, 배수관계를 확인 할 경우에도 10(0.1s) 이하의 경우는 허용 가능한 오차로 간주하였다. 즉 [그림 6]의 (25, 3)에서 3은 25의 선행음에 연결, 첫 음이 28의 길이로 발생된 것으로 간주하였다. [그림 6]에서 첫 번째 열의 두 사각형들은 처음 발생음 $T_0(28)$ 를 기준으로 음표 조합 허용 오차인 10(0.1s) 범위 내에서의 개별 음들을 조합하였을 때 T_1 까지 조합이 가능함을 보여준다. 이때의 조합 오차는 $a=|b-c|=5$ 이며, 누적인 음의 길이는 $S_1(61)$ 이다.

이와 같이 첫 번째 조합이 시도 된 후 두 번째 조합은 [그림 5]의 b의 과정과 같이 T_0 의 다음 음인 T_1 까지의 누적 합계인 $S_1(61)$ 을 기준으로 진행된다. $T_2(27)$ 와 $T_3(30)$ 가 조합 오차 범위 4 이내에서 조합된다. 동일한 방법으로 세 번째 조합은 두 번째 조합의 T_1 다음 음인 T_2 까지의 누적 합계인 $S_2(88)$ 을 기준으로 진행되며, T_3, T_4, T_5 가 오차 범위 7 이내에서 조합된다. 마지막으로 [그림 6]의 네 번째 열은 T_3 까지의 누적인 음의 길이 $S_3(118)$ 을 기준으로 $T_7(29)$ 까지 오차 범위 내에서의 조합을 수행한다.

이상과 같이 4번의 음표 조합 과정이 끝나게 되면 각 조합 과정에서 발생한 S_0, S_1, S_2, S_3 에 대해 배수 관계를 조사한다. [그림 6]에서 확인할 수 있듯이 S_1, S_2, S_3 은 S_0 에 대해 각각 2,3,4배의 배수 관계가 존재함을 알 수 있으며, S_3 은 다시 S_1 에 대해 2배의 배수 관계를 가지고 있음을 발견 할 수 있다. 징검다리 배수에 대한 정의 $S_k/S_i = l_{ij}l_{jk} = l_{ik}$ 에 따라 $S_3/S_0 = l_{01}l_{13} = l_{03}$ 이 성립하므로 징검다리 배수 위치는 61이 되며 마디가 존재하는 지점은 118(1180ms) 위치임을 알 수 있다.

2. 규칙2 - 징검다리 배수가 존재 하지 않을 경우

2.1 하나의 배수 관계만이 존재할 경우

만약 징검다리 배수가 없고 단지 하나의 배수 관계만이 존재 한다면, 배수의 마지막 종착부분이 마디의 위치이다.

마디 찾기를 위한 두 번째 규칙은 [그림 8]과 같이 징검다리 배수관계가 없고 음표 조합에 대한 배수관계가 하나만 존재할 경우이다. [그림 8]은 규칙 2에 대한 4번의 조합 과정을 보여주는 것으로 정확한 마디 위치는 2분 음표 조합 길이에 해당하는 위치임을 알 수 있다.

(a) 실제 악보

(b) 이상적인 음표 조합 과정

그림 7. 규칙 2에 대한 이상적인 음표 조합 과정

규칙1과, 규칙 2.1에 속하지 않으며, 복수개의 배수관계를 가지는 경우는 4번까지의 조합 횟수를 추가적으로 확장한다. 이는 음악적인 음표들에 대한 조합은 반드시 징검다리 배수관계가 나타난다는 음악적 지식에 기반 한 것이다.

(a) 실제 악보

(b) 이상적인 음표 조합 과정

그림 9. 규칙 3에 대한 이상적인 음표조합 과정

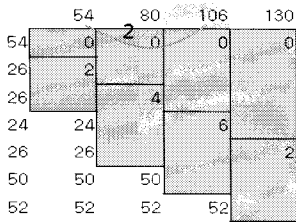
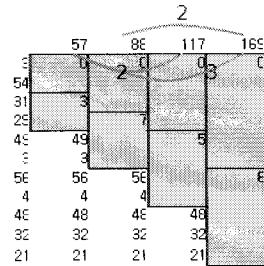


그림 8. 그림 7 노래의 실제 발생에 대한 조합과정

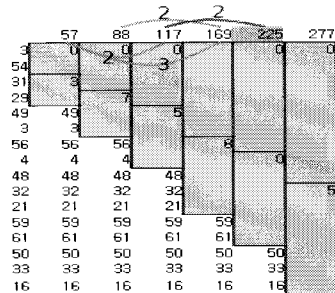
이는 이상적인 음표 조합 과정인 [그림 7]을 통해 확인 할 수 있다. [그림 7]에서 2분 음표를 기준으로 조합을 시도했을 때 이상적인 마디 위치가 발견됨을 알 수 있고 [그림 8]과 같이 실제 발생된 데이터에 대해서 배수 관계를 조사 했을 때 54와 106 사이의 배수의 배수 관계 밖에는 존재하지 않아 규칙에 의하여 배수 관계의 마지막 종착 부분인 106이 마디의 위치임을 알 수 있다.

2.2 두 개 이상의 배수 관계가 존재할 경우

만약 징검다리 배수가 없고 단지 두 개 이상의 배수 관계가 존재 한다면, 최대 조합 횟수를 확장하여 징검다리 배수를 찾는다.



(a) 첫 번째 조합 과정



(b) 확장된 조합 결과

그림 10. 그림 9 노래의 실제 발생에 대한 조합 과정

[그림 10]은 [그림 9]의 원 곡에 대해 발생한 초기 조합 결과를 보여주는 것으로 [그림 10]의 (a)는 징검다리

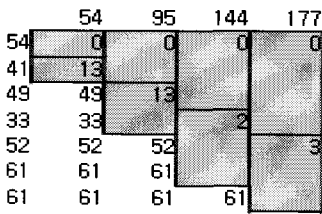
배수관계가 존재 하지 않고 3개의 배수관계가 있음을 알 수 있다. 실제로 이 정보만 가지고서는 마디의 정확한 위치를 알 수 없기 때문에 [그림 10]의 (b)와 같이 4회 조합 이후에 추가적인 조합을 시도한다. [그림 10]의 (b)에서 다섯째 열의 225 위치는 57, 117의 배수관계를 만족하므로 장검다리 위치는 117이며 마디위치는 225에 해당한다. 즉 음표를 추가 확장하여 조합을 시도했을 때 규칙1을 적용할 수 있는 장검다리 배수가 존재하게 되어 마디 위치를 정확히 찾을 수 있다.

3. 규칙3

만약 어떠한 배수관계도 존재하지 않는다면, 조합의 시작 위치를 조절하여 배수관계가 존재하는 가장 먼 조합 시작 위치를 찾는다.



(a) 실제 악보



(b) 음표 조합 결과

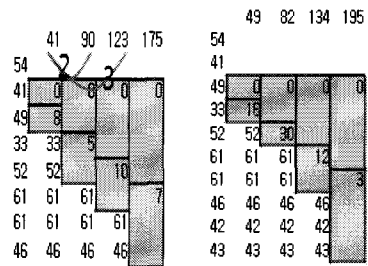
그림 11. 실제 악보와 발생된 노래의 음표 조합 결과

규칙3은 이상적인 마디 조합 과정에는 존재하지 않고 오직 사람이 악보와 다르게 노래하기 때문에 나타난다. 즉 실제 악보에는 음표들이 존재하며 이러한 음표들의 조합은 반드시 배수관계를 가지지만 사람의 노래는 가변적이기 때문에 어떠한 배수관계도 존재하지 않을 수 있다.

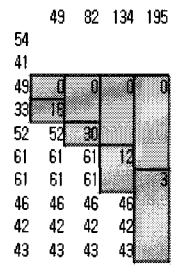
[그림 11]은 규칙3의 필요성을 묘사한 것으로 [그림 11]의 (a)노래에 대해 발생했을 때 어떠한 배수관계도 포함하고 있지 않고 있다. 이러한 경우는 최초 조합 시작 음의 길이가 지나치게 길거나 짧을 때 나타나기도 한다.

본 논문에서는 이러한 형태를 가지는 조합 결과에 대

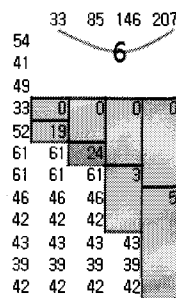
해서는 최초 조합 시작 위치를 변경시켜서 배수관계를 찾고, 해당 배수관계를 가운데 가장 멀리 위치한 조합 시작 위치를 마디 위치로 설정하였다. 이는 발생의 형태에 따른 모호함으로 인해 최초 조합 시작 위치로는 마디를 찾기 어렵다고 판단하고 조합 시작 위치를 바꾸어 배수관계를 얻기 위한 시도이다. 또한 이렇게 얻어진 배수관계들도 조합 시작 위치에 따라 여러 형태로 존재 할 수 있는데, 배수관계가 존재하는 가장 마지막 조합 시작 위치를 마디로 보는 이유는 최초 발생된 조합 시작 위치 음으로부터 최대한 멀리 떨어져서 조합된 결과가 이상적이라 판단했기 때문이다.



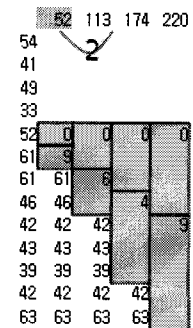
(a)



(b)



(c)



(d)

그림 12. 규칙4를 적용한 네 번째 발생 노래에 대한 조합 과정:(a)첫 번째 조합 시작 위치 변경 결과, (b)두 번째 조합 시작 위치 변경 결과, (c)세 번째 조합 시작 위치 변경 결과, (d)네 번째 조합 시작 위치 변경 결과

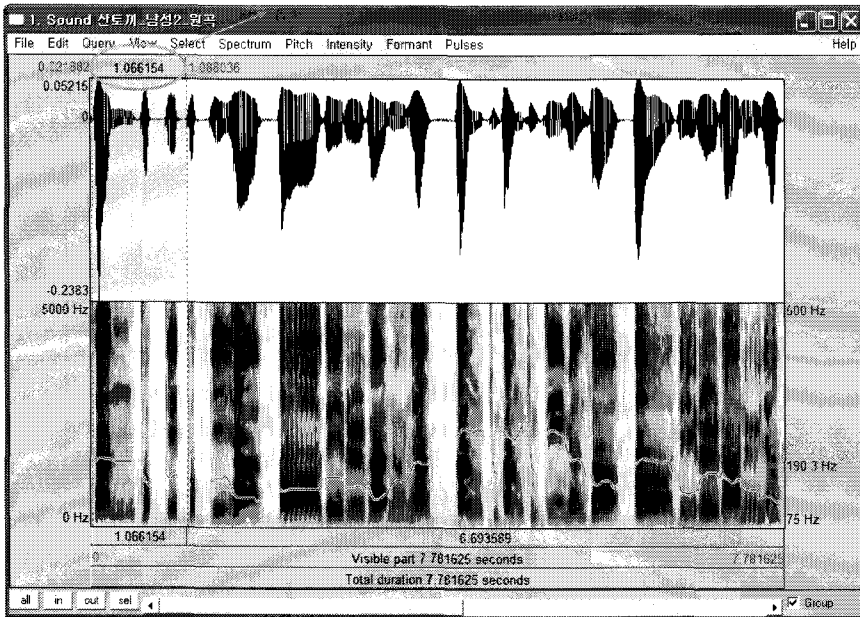


그림 13. praat 소프트웨어를 이용한 마디 검출 위치 확인

이것은 첫 번째 마디보다는 이후의 마디를 먼저 찾고 이를 기반 하여 첫 번째 마디 위치를 찾는 과정으로 볼 수 있다.

[그림 12]는 이러한 과정을 보여주는 것으로 조합 시작 위치가 [그림 11]의 (b)에서의 조합 시작 위치인 54에서 41로 변경했을 때 나타나는 배수관계를 확인할 수 있다. 최대 4번의 조합 시작 위치를 변경하였을 때 [그림 12]의 (d)와 같이 가장 멀리 시작한 조합 위치에서 마디를 검출하였음을 보여준다.

IV. 실험 및 고찰

본 논문에서 실험에 사용된 노래는 10곡의 동요와 애국가, 가요1곡(어머나)로서 총 12개의 곡이 사용되었다. 각 노래들은 일체의 악기 소리 없이 오직 사람의 목소리만으로 구성되어 있다. 좀 더 원활한 데이터 수집을 위해 피실 협자들(남성, 여성 각각 2명씩)은 각자 개인 컴퓨터 앞에서 마이크를 이용하여 해당 노래를 녹음 하였고, 수집된 노래들은 praat[13] 소프트웨어를 이용하여 pitch정보만을 추출하여 실험에 사용하였다.





[표 2]는 실험에 사용된 12개 곡에 대한 분석 결과를 나타낸 것으로 3장에서 기술한 마디 찾기 규칙들의 종류와 찾아진 박자 및 빠르기 정보를 보여준다.

[그림 13]은 [표 2]의 song1(man2)에 대한 praat 소프트웨어를 이용한 음성 파형의 분포를 나타낸 것으로, [그림 13]의 a는 본 논문에서 제안한 규칙을 이용하여 찾아진 마디위치를 (106) 표시한 것이다. praat 소프트웨어는 [그림 13]의 a 와 같이 선택한 영역에 해당하는 부분을 재생해 주는 기능을 제공해 주기 때문에 선택한 영역이 실제 악보상의 어느 위치까지 인지를 쉽게 알 수 있다. 이를 통해 본 논문에서 제시한 규칙을 통해 찾아진 마디의 위치가 실제 악보상의 어느 부분에 해당하는지를 [표 2]와 같이 표시하여 그 정확성의 유무를 판별하였다.

실험 결과 본 논문에서 제안한 마디 찾기 위한 조합 방법을 이용하면 song1, song 3, song 4, song 9와 같이 비록 동일한 곡에 대해 서로 다른 사람이 발성한 노래라 할지라도 마디의 위치를 정확하게 찾고 있음을 확인할 수 있다.

본 논문에서는 변박 자가 없다고 가정하고 곡의 빠르기가 처음부터 끝까지 동일한 것으로 가정했기 때문에

표 2. 실험 결과

| | Singer Type | Rule Type | Measure Position(ms) | beat Type | tempo | |
|--------|-------------|-----------|----------------------|---|-------|-----|
| Song1 | man1 | 2 | 101 |  | 3 | 237 |
| | man2 | 2 | 106 |  | 4 | 226 |
| Song2 | woman1 | 1 | 118 |  | 3 | 203 |
| Song3 | man1 | 4 | 299 |  | 4 | 80 |
| | man2 | 2 | 305 |  | 4 | 78 |
| Song4 | man1 | 3 | 152 |  | 3 | 157 |
| | woman1 | 1 | 150 |  | 3 | 160 |
| Song5 | woman1 | 2 | 265 |  | 3 | 90 |
| Song6 | woman2 | 3 | 129 |  | 4 | 186 |
| Song7 | woman1 | 4 | 194 |  | 4 | 123 |
| Song8 | woman1 | 3 | 117 |  | 4 | 205 |
| Song9 | man1 | 1 | 203 |  | 4 | 118 |
| | woman1 | 1 | 207 |  | 4 | 115 |
| Song10 | woman1 | 1 | 209 |  | 4 | 114 |
| Song11 | woman1 | 2 | 210 |  | 3 | 114 |
| Song12 | man1 | 4 | 177 |  | 3 | 135 |

첫 번째 마디의 박자 정보와 빠르기가 곡 전체를 대표하며, 6/8 박자 및 2/4박자는 각각 3/4, 4/4박자의 확장 및 축소 영역에 해당하기 때문에 3박자 계통과 4박자 계통에 대한 관별만을 시도하였다.

[표 2]의 빠르기에 대한 평가는 만약 우리가 첫 번째 마디의 위치를 정확하게 찾았다면 이는 실제 발생된 노래에서의 첫 번째 마디의 재생 위치, 즉 마디의 재생 시간이 일치하게 되므로 음악 이론상 본 논문에서 찾는 빠르기가 실제 발생된 노래의 빠르기가 되는 것으로 대

체하였다.

song2, song5에서는 마디의 위치가 실제 마디 위치보다 1/2 지점에 위치하고 있지만 실제 박자의 중간 지점 이므로 빠르기 정보 또한 달라지는 부분은 없다. 실제로 NWC 등의 작곡 프로그램을 이용하여 song2, song5에 대해 빠르기를 지정하여 연주했을 때 실제 발생된 노래와 동일한 빠르기로 연주하고 있음을 확인하였다.

[표 2]에서 song11은 마디의 위치를 잘못 찾은 경우에 해당하는 것으로 [그림 14]에서 알 수 있듯이 4개음



(a) 실제 악보

| | | | | |
|----|----|-----|-----|-----|
| | 60 | 103 | 157 | 210 |
| 60 | 0 | 0 | 0 | 0 |
| 37 | 13 | | 2 | |
| 6 | | | | |
| 54 | 54 | 4 | | |
| 53 | 53 | | 8 | |
| 44 | 44 | 44 | | 114 |
| 52 | 52 | 52 | | |

(a) 음표 조합 결과

그림 14. song11의 두 번째 규칙을 적용한 조합 결과

의 조합만을 시도하기 때문에 이상적인 마디 위치를 찾고 있지 못하고 있음을 알 수 있다. 특별히 [그림 11]과 같이 한 마디 안에 짧은 비트의 동일한 길이를 가지는 음표들로 구성된 노래들에 대해서는 본 논문에서 제안한 규칙이 적용되지 않고 있다. 조합 음의 수를 확장하면 간단히 해결 할 수 있지만 다른 규칙과의 충돌로 인해 이상적인 방법이라고는 말할 수 없다. song11과 유사한 노래들을 비롯한 다양한 조합 형태를 가지는 곡들에 대한 추가적인 규칙의 연구가 필요하다고 말할 수 있다.

song1-1, song2, song12의 곡들에 대한 박자 검출은 모두 3/4 박자 유형으로 판별되었다. 이는 비록 마디의 위치가 정확하다 할지라도 발생된 음표들의 길이가 한 마디를 4박자로 완벽하게 채울 수 없기 때문에 발생하는 것으로 단순히 음표 타입 길이에 대한 합계만으로는 박자를 판단하기에는 부족한 부분이 있음을 알 수 있다. 좀 더 정확한 박자 정보 검출을 위해서는 마디 내에 분포하는 음표들의 리듬 정보 등을 결합하여 활용해야 할 것으로 판단된다.

V. 결론

본 논문에서는 사람이 발성한 개별 음들을 좀 더 정확한 분할 위치를 찾아내기 위해 샘플링된 피치 값들의

이동 표준 편차의 정보와 GMM을 활용하였다. 또한 자동 음악 채보 시스템의 중요 사전 정보에 해당하는 곡의 박자 및 빠르기를 찾아내기 위해 마디 정보의 필요성을 언급하고 이를 이용하여 정확한 마디 위치를 찾기 위한 3가지 규칙들을 제안하였다.

이는 기존의 자동 음악 채보 시스템들이 가지고 있었던 곡의 박자와 빠르기를 사전에 파라미터 형태로 입력해야만 했던 단점들을 보완 해 줄 수 있을 것으로 기대된다.

본 논문에서는 3박자와 4박자 계통에 대한 박자 인식을 시도하였다. 이는 보편적으로 3박자와 4박자 계통의 노래들이 많기 때문이다. 좀 더 전문화되고 다양한 분야의 노래들에 대한 박자 검출을 위해서는 추가적인 음악적 지식을 기초로 한 규칙의 발견이 필요할 것으로 판단되며, 못 갖춘마디를 비롯한 본 논문에서 실험에 사용된 단순한 형태의 음표 구성이 아닌 복잡한 음표가 사용되는 곡들에 대한 추가적인 음표 조합 연구가 필요할 것으로 판단된다.

참고 문헌

- [1] “2005 음악산업백서,” 문화관광부.
- [2] George Tzanetakis and Perry Cook, “Musical Genre Classification of Audio Signals,” IEEE Transaction on Speech and Audio Processing, Vol.10, No.5, 2002(7).
- [3] 이윤주, 문성빈, “내용기반 음악정보 검색시스템을 위한 이용자 중심의 질의 인터페이스 설계에 관한 연구”, 정보관리학회, 제23권 2호, pp 5-19, 2006.
- [4] <http://monstar.fm>
- [5] <http://www.creivedectors.com>
- [6] <http://www.recognisoft.com>
- [7] Matija Marolt and Marko Privosnik, “SONIC : A System for Transcription of Piano Music,” Cybernetics and Systems, 2002.
- [8] Kashino, Nakadai, Kinoshita and Tanaka,

"Application of Bayesian probability network to music scene analysis", Computational auditory scene analysis, pp 115-137, 1998.

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

[10] C.Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking", Proc.Computer Vision and Pattern Recognition 99, Colorado, 1999(6).

[11] 형아영, 이준환, "SIDE를 이용한 자동 음악 채보 시스템", 정보처리학회, 제16_B권 2호, pp 141-150, 2009.

[12] <http://www.praat.org>

이준환(Joon-Whoan Lee)

정회원



- 1980년 : 한양대학교 전자공학과(공학사)
 - 1982년 : 한국과학기술원 전자공학과(공학석사)
 - 1982년 ~ 1985년 : 전북대학교 전자공학과 조교
 - 1985년 ~ 1987년 : 전북대학교 전자공학과 전임강사
 - 1990년 : 미주리대학 전산학과(공학박사)
 - 1990년 ~ 현재 : 전북대학교 전자정보공학부 교수
- <관심분야> : 내용기반 영상검색, 멀티미디어

저자 소개

박은중(En-Jong Park)

정회원



- 2002년 2월 : 군산대학교 컴퓨터과학과(공학사)
- 2005년 2월 : 전북대학교 전자공학과(공학석사)
- 2008년 8월 : 전북대학교 영상공학과(공학박사)

▪ 2009년 8월 ~ 현재 : 한국전자통신 연구원
<관심분야> : 감성기반 영상검색, 멀티미디어

신송이(Song-Yi Shin)

준회원



- 2008년 2월 : 전북대학교 전자정보공학부(공학사)
- 2008년 3월 ~ 현재 : 전북대학교 컴퓨터공학과(공학석사 재학 중)

<관심분야> : 감성기반 음악검색, 멀티미디어