

# 판소리 자동채보를 위한 구조분석 알고리즘

## Structural Analysis Algorithm for Automatic Transcription 'Pansori'

주영호\*, 김준철\*\*, 서경숙\*\*\*, 이준환\*

전북대학교 전자정보공학부 컴퓨터공학\*, 서남대학교 전자공학과\*\*, 전북도립국악원\*\*\*

Young-Ho Ju(nightmute@jbnu.ac.kr)\*, Joon-Cheol Kim(kjc1029@daum.net)\*\*,  
Kyoung-Suk Seo(skidtom123@hanmail.net)\*\*\*, Joon-Whoan Lee(chlee@jbnu.ac.kr)\*

### 요약

서양 음악의 경우 자동채보와 내용기반 음악검색을 위한 음악 정보 분석연구가 활발하게 진행되고 있다. 그러나 한국 전통음악에서는 유사한 연구사례를 찾아보기 어렵다. 본 논문에서는 한국의 전통음악인 판소리 구조를 자동으로 분석하기 위한 알고리즘들을 제안한다. 제안된 알고리즘은 음성과 비음성의 시간 간격 비율을 이용하여 '소리' 부분과 '아니리' 부분을 자동으로 구분한다. 뿐만 아니라 알고리즘은 '장단'이라 칭하는 리듬을 템플릿 이용한 다수결 결정 방법으로 강건하게 구분한다. 또한 알고리즘은 칼만 필터를 이용하여 '소리' 부분의 마디 지점을 검지해낸다. 본 논문에서 제안된 알고리즘들은 판소리 샘플들에서 양호하게 동작하였으며 자동채보의 전단계의 구조분석에 유용할 수 있다.

■ 중심어 : | 판소리 구조분석 | 장단 | 자동채보 | 템플릿 유사도 | 칼만필터 |

### Abstract

For western music there has been a volume of researches on music information analysis for automatic transcription or content-based music retrieval. But it is hard to find the similar research on Korean traditional music. In this paper we propose several algorithms to automatically analyze the structure of Korean traditional music 'Pansori'. The proposed algorithm automatically distinguishes between the 'sound' part and 'speech' part which are named 'sori' and 'aniri', respectively, using the ratio of phonetic and pause time intervals. For rhythm called 'jangdan' classification the algorithm makes the robust decision using the majority voting process based on template matching. Also an algorithm is suggested to detect the bar positions in the 'sori' part based on Kalman filter. Every proposed algorithm in the paper works so well enough for the sample music sources of 'Pansori' that the results may be used to automatically transcribe the 'Pansori'.

■ keyword : | Pansori Structure Analysis | Jangdan | Automatic Transcription | Template Matching | Kalman Filter |

## 1. 서론

인터넷을 통한 디지털 음원 서비스가 보편화 되면서 다양한 음원 분석과 내용기반 음원검색 연구 역시

ISMIR(International Society for Music Information Retrieval) 등을 중심으로 발전해가고 있다. 매년 열리는 ISMIR에서는 MIREX(Music Information Retrieval

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

접수일자 : 2013년 10월 14일

접수일자 : 2014년 02월 10일

심사완료일 : 2014년 02월 11일

교신저자 : 이준환, e-mail : chlee@jbnu.ac.kr

Evaluation eXchange)라는 경쟁을 통해 기술 및 데이터 교류가 진행되는데[1], 서양음악을 대상으로 장르 분별, 분위기 분별, 온셋(onset) 검출, 건(key) 검출, 템포 검출 등 21개의 다양한 분야의 경쟁이 진행된다.

이러한 서양음악에 대한 정보공학적 분석 및 활용은 미국, 유럽, 일본 등 선진국에서는 이미 오래전부터 신호처리, 인공지능, 음악 분야의 학자들이 공동으로 연구에 참여하여 활발하게 진행되어 왔으며[2-5], 국내에서도 최근에 융합연구들이 보편화됨에 따라 음악의 자동채보, 음악의 장르분별, 음악의 감성분별 및 내용기반 음원 검색의 연구들이 진행되고 있다[6-10].

본 논문에서는 CD나 공연 실황이 녹음된 음원으로부터 자동으로 판소리를 채보하는 시스템에 필수적인 판소리 구조분석 알고리즘들을 제안하였다. 음원을 분석하는 방법은 온라인 또는 오프라인으로 분석하는 방법이 있다. 본 논문에서 제안하는 방법은 오프라인 분석으로, 판소리를 구성하는 소리와 아니리 부분을 구분하고 다시 소리부분에서 장단을 구별하여 장단을 구성하는 마디를 찾아내는 알고리즘들을 제안한다.

제안하는 방법은 실험 샘플음원에 대해 아니리와 소리를 만족스럽게 구별하였으며, 진양조, 중모리, 중중모리, 자진모리 4개의 장단을 정확하게 구분하였다. 이 장단 구분과정에서는 북소리에 해당하는 ‘합’과 ‘궁’(진양조의 경우)이라 부르는 부분의 장단에 따른 특징을 이용하며, 다수결 의사결정 알고리즘을 통해 장단구별의 정확성을 제고하였다. 또한 본 논문에서는 판소리 장단의 템포 변화와 마디 지속시간을 추정하기 위해 칼만 필터(Kalman Filter)를 이용한 알고리즘을 제안하였으며, 근사적으로 마디 길이의  $\pm 5\%$  범위에서 마디길이를 정확하게 예측하였다.

2003년 11월 유네스코 무형문화유산으로 등재된 판소리는 민속음악이며 서민들 사이 유행하던 비정형화된 음악이었기 때문에 체계화되어 있지 못하다. 본 논문에서의 장단의 자동분별 결과는 장단에 따른 분위기를 분별하는데 유용하며 마디의 구분은 유파나 창자에 따른 템포 변화를 분석하는데 사용할 수 있으며, 또는 서양음악과 같은 기보 방법을 적용할 경우의 음길이를 부호화하는데 유용하게 활용될 수 있다. 또한 아니리와

소리, 장단과 마디들은 아니리 부분의 경우 음성인식을 통해 그 내용을 찾아내는데 활용할 수 있으며, 소리 부분은 피치를 추출하여 음정을 부호화하는데 활용할 수 있다.

기존의 음악의 자동채보 및 분석에 관련된 연구들은 전술한 바와 같이 ISMIR을 중심으로 다양한 연구들이 진행되어 왔으며[4][5], 국내의 경우도 몇몇 대학에서 시도한 바 있다[7][10]. 그러나 이들 연구들은 모두 서양음악이나 단음형태의 우리 동요 등을 대상으로 하였으며, 우리 전통음악에 대한 시도는 극히 제한적이었다. 예로써 박형순은 녹음실에서 창자의 목소리를 직접 녹음하고 음성 시각기를 이용하여 사람 발성의 음향학적 형태를 구별하여 채보하는 방법을 제시하였다. 하지만 이 방법은 사람이 직접 듣고 판단하며 일일이 녹음 구간을 나누어 사용하고 있어 진정한 자동채보 또는 분석 방법으로 간주되기 어렵다[11]. 따라서 본 논문에서 제안된 방법들은 우리 전통음악의 자동채보를 위한 시발점으로서의 구조분석이라는 측면에서 의의가 있다.

본 논문의 2절에서는 본 논문에서 분석 대상으로 하는 판소리의 구조를 살펴보고, 3절에서는 제안된 알고리즘을 기술하며 4절에서는 제안된 알고리즘의 실험결과를 살펴보았으며 5절에서는 본 논문의 의의와 제안된 구조분석 알고리즘의 활용방안을 제시하였다.

## II. 판소리의 구조

판소리는 소리꾼이 병풍을 두르고 돛자리를 펼친 마당이나 공연장에서 긴 이야기를 몸짓을 섞어가며 고수의 북 반주에 맞춰 노래하는 한국 전통음악이다. 판소리의 주요한 3대 요소는 소리꾼, 고수, 추임새를 같이하며 참여하는 청중이라 할 수 있다[12]. 본 절에서는 논문에서 목표로 하는 구조분석에 필요한 소리와 아니리 장단 등을 간략하게 기술한다.

### 1. 소리와 아니리

판소리는 단순히 이야기 줄거리를 나열한 노래부분만으로 이루어진 것이 아니라 이야기를 이어가며 창의

력을 통해 다채롭게 표현한다. 이를 위해 노래하는 소리(노래하는 것)외에도 아니리(소리를 하다가 이야기의 장면을 대화식으로 설명하는 것)와 발림(소리를 하는 가운데 동작이나 손짓, 몸짓으로 감정을 표현하는 것)을 활용한다.

## 2. 장단

장단이란 각기 리듬 꼴이 다른 작은 단위가 여러 개 모여서 한 리듬 패턴을 형성한 것이라고 말할 수 있다. 다양한 감정 표출을 요구하는 판소리에서 장단은 음악의 흐름을 통제하는 기본 요소로 비중이 크다.

오늘날 판소리에 쓰이고 있는 장단은 진양조, 중모리, 중중모리, 자진모리, 휘모리, 하모리, 엇중모리를 포함한 일곱 가지 종류로 구별되고, 속도나 박자의 분할 방법에 따라 세마치(빠른 진양), 단중모리(빠른 중모리) 같은 세분된 용어들이 사용된다. 다음에는 본 논문에서 대상으로 한 진양조, 중모리, 중중모리, 자진모리 장단의 기본 꼴을 설명한다.

### 2.1 진양조

진양조는 판소리에서 가장 느린 장단으로 주로 한가한 장면, 조용하고 편안한 장면, 슬퍼서 탄식하는 장면, 서정적인 대목 등에 쓰인다. 진양조는 3분박 느린 6박자로, 이 6박자를 '한 각'이라고 한다. 흔히 "밀고" "달고" "땃고" "풀고"하는 4각을 주기로 변주하기 때문에 진양 장단을 4각 24박자라고도 한다. 이들 24박의 시작에는 '합', 그 외의 각의 시작에는 '궁'이라는 북소리가 수반되며 네 개의 각을 각각 다르게 치지만, 단순히 시간적 배분으로만 생각하면 똑같은 여섯 개의 마디가 네 번 반복되는 것으로 볼 수 있다.

### 2.2 중모리

중모리는 보통 속도의 장단으로 판소리 장단의 기본이라고 할 수 있다. 서술적인 대목, 어느 사연을 평탄한 어조로 말하는 대목, 서정적인 대목같이 태연스럽고 안정된 분위기에 주로 쓰인다. 우리 음악이 대부분 3분박인데 비해 중모리는 예외적으로 2분박에다 보통 빠르기인 12박자이다. 장단은 합이라는 북소리에 의해 12박

자가 시작되기 때문에 마디는 합에 의해 구별된다.

### 2.3 중중모리

중모리의 속도가 빨라져서 이루어진 장단으로 조금 빠른 장단 주로 흥겨운 대목, 흥분하고 통곡하는 대목 등에 쓰인다. 3분박, 조금 느린 4박자로 친다. 장단은 중모리와 같이 합이라는 북소리에 의해 12 박자가 시작되기 때문에 마디는 합에 의해 구별된다.

### 2.4 자진모리

중중모리가 빨라져서 된 장단으로 주로 어떤 상황을 차례로 늘어놓거나 박력 있고 긴박하며 명랑한 분위기에 쓰이며 대개 극적인 장면이 많다. 일반적으로 3분박 빠른 4박자로 친다. 속도 때문에 잔가락 변주나 분할이 불가능해 저서 단조로운 리듬 꼴이 된다. 장단은 중모리, 중중모리와 같이 합이라는 북소리에 의해 12 박자가 시작되기 때문에 마디는 합에 의해 구별된다.

## III. 제안된 판소리 구조분석 알고리즘

본 절에서는 자동 판소리 채보를 위한 구조분석 과정을 개괄적으로 설명한 후, 입력된 판소리 음원 신호가 시스템에서 처리 분석 되는 순서에 따라 각 단계별로 제안된 알고리즘에 대하여 설명한다.

### 1. 판소리 채보를 위한 구조분석 과정

본 논문의 궁극적인 목적은 공연 실황을 녹음한 음원을 자동으로 분석하여 채보하는 시스템을 개발하기 위한 것으로서, 해당 음원의 소리꾼의 발성 정보와 고수의 소리 정보를 이용한다.

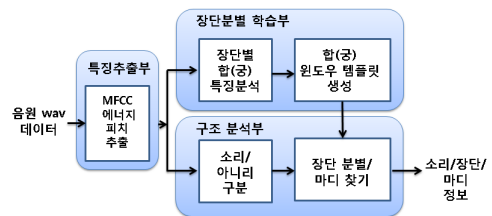


그림 1. 판소리 자동 구조분석 과정

[그림 1]은 본 논문에서 제안한 자동 구조분석 과정을 보여 주고 있다. 분석과정에서는 수집된 판소리를 컴퓨터에서 처리할 수 있는 wav파일(.wav) 형태로 입력하며, 훈련 집합과 검증 집합으로 나누어 처리한다.

### 1.1 특징 추출부

우선 훈련 및 검증 집합의 모든 음원에서 합과 공에 해당하는 북소리와 음성부분을 찾기 위해 피치(pitch), 에너지, MFCC(Mel-Frequency Cepstral Coefficients) 등의 특징을 추출한다.

피치정보는 음높이를 표현하며 단일음의 주파수와 관계된 지각적인 특성으로 AMDF (Average Magnitude Difference Function)

$$D_x(m) = \frac{1}{N} \sum_{n=0}^{N-1-m} |x(n) - x(n+m)|, \quad (1)$$

$$0 \leq m \leq M_0$$

를 이용하여 추출하였다. 식 (1)은 샘플링된 음성신호  $x(n)$ 과 이 위치에서  $m$ 만큼 이동된 음성신호  $x(n+m)$ 의 차의 절대 값의 평균으로 본 연구에서 샘플링 주파수는 16000Hz이고 프레임 크기  $N$ 은 320, 창(window) 함수의 넓이는 160으로 설정해 매 0.01초마다 피치정보를 추출하였다[13]. 피치 정보는 소리와 아니리 구분을 위해 음성(phonetic)과 비 음성(pause) 부분을 판별하는데 이용되고, 자동채보에서는 음높이를 부호화하는 기본 정보를 제공하며, 동일 피치의 기간은 음길이를 부호화 하는데 활용된다.

또한 음원의 에너지란 시간에 따른 음의 세기(amplitude)를 보여주는 특징으로

$$E_n = \frac{1}{N} \sum_m [x(m)w(n-m)]^2 \quad (2)$$

로 표현되며, 여기서  $w(n)$ 은 창 함수를 표현한다[14]. 판소리 구조분석에서 에너지 특징은 고수의 북소리의 강약을 검출하거나 소리꾼 발성의 시작 부분 또는 강조 부분 등을 찾아내는데 유용하다.

MFCC 특징은 주파수 분포 중에서 인간의 청각시스템에 민감한 부분을 포착하는 13차원의 벡터로 소리 분류에 사용되는 대표적인 특징이다[15]. 본 논문의 판소

리 구조 분석에서 MFCC는 서로 다른 장단에서의 합(궁) 소리의 특징을 표현하는데 유용하다.

### 1.2 아니리/소리 구분

전술한 바와 같이 판소리 음원의 경우 소리 부분뿐만 아니라 아니리가 포함되어 있다. 아니리 부분의 채보는 무의미 하므로(음성인식의 대상이 되어야 함), 우선적으로 음원을 소리와 아니리 구분을 하고 소리 부분에 대해서만 채보 과정이 필요하다. 제안된 방법의 아니리와 소리구분에서는 의미있는 피치 정보의 유무에 따른 음성구간과 비 음성 구간의 판별 결과를 이용하였다. 즉 아니리 부분은 사람의 말소리와 같아 비 음성 구간이 음성구간에 비해 많고 비 음성 구간이 짧게 끊어지는 특징을 가진 반면, 소리 부분은 비 음성 구간보다 음성구간의 길이가 길고 비율도 높다. 또한 대부분의 소리시작 부분에서 소리꾼이 크게 호흡을 하기 때문에, 약 0.3초 이상의 비 음성구간이 존재 한다. [그림 2]는 아니리 부분과 소리의 시작 지점의 특징을 보여주며, 본 논문에서 제안하는 아니리와 소리를 구분하는 알고리즘은 [표 1]과 같다.

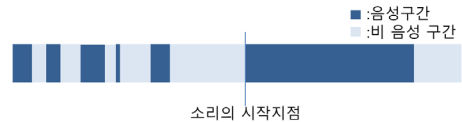


그림 2. 아니리 부분과 소리의 시작 지점의 특징

표 1. 음원의 소리 부분 추출 알고리즘

Step 1	음원내의 피치 정보 유무에 따라 음성 구간과 비 음성 구간을 구분
Step 2	만약 현재 비음성 구간의 길이 > 0.3초 이고 이전 10개의 구간에서 (음성 구간/비음성 구간) < 1.0 이하였고 다음 10개의 구간에서 (음성 구간/비음성 구간) > 1.0 이상일 때 현재의 위치부터 소리의 시작 부분으로 판별, Step 3을 진행
Step 3	만약 현재 비음성 구간의 길이 > 0.3초 이고 이전 10개의 구간에서 (음성 구간/비음성 구간) > 1.0 이하였고 다음 10개의 구간에서 (음성 구간/비음성 구간) < 1.0 이상일 때 현재의 위치부터 아니리의 시작 부분으로 판별, Step 2를 진행
Step 4	음원의 끝까지 Step 2, Step 3을 반복

1.3 장단분별 학습부

제안된 방법에서는 훈련 집합에 속해있는 판소리 음원들의 소리 부분에서 전문가가 미리 추정해낸 합과 궁(진양조의 경우)의 위치를 이용한다. 한편 전술한 바와 같이 음향의 특징을 나타내는 13차원 벡터인 MFCC, 에너지, 피치 등 3가지 특징을 시간 축 상에 늘어놓으면 한 시간 단면에서 15차원의 벡터가 되며, 이를 미리 추정된 합(궁)의 전후에서 매 0.01초 단위로 취득하면 [그림 3]의 한 합(궁) 부분에서와 같은 모습의 15차원 벡터 열이 생성된다. 따라서 훈련 집합에 속하는 장단이 알려진 소리마다 합(궁)부분(실제로는 사람에 의해 근사적으로 추정된 합(궁) 전후에서의 에너지 값의 국부 최대 지점)을 전후해서, 이들을 추출하고 이들을 평균하면 각 장단의 합(궁)분별을 위한 15차원 벡터 열 형태의 윈도우 템플릿을 구성할 수 있다.

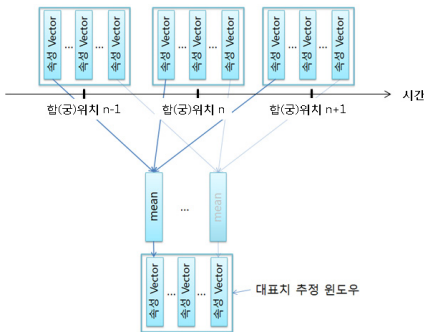


그림 3. 합(궁) 분별 윈도우 평균템플릿 생성 과정

실제적으로 합(궁)부분에서 북소리의 지속시간은 장단마다 다르고 이를 최적으로 검출하기 위한 윈도우 템플릿의 지속시간 또한 달라야 한다. 따라서 제안된 방법에서는 각 장단의 최적의 윈도우와 지속시간을 [표 2]와 같이 예측하였다. 즉 100ms부터 2000ms까지 100ms 단위로 [그림 3]에서와 같이 추출한 윈도우 템플릿의 기간을 다르게 하여, 사람에 의해 근사적으로 추정된 합(궁)위치 주변의 에너지의 국부 최대치 부분을 찾아 합(궁)위치라고 가정할 때, 이들 위치들에서 NCC(Normalized Cross Correlation)가 최대 평균값을 갖는 템플릿을 최적 윈도우 템플릿과 지속시간으로 간주하였다.

표 2. 장단분별 합(궁) 분별 최적 윈도우 템플릿 산출 과정

Step 1	각각의 장단에 대해 모든 추정된 합(궁) 위치에서 에너지의 국부 최대치 부분을 NLMS에 의해 탐색함.
Step 2	각각의 장단에 대해 Step 1의 각각의 에너지 국부 최대치 부분을 전후로 0.1초부터 2초까지 0.1초씩 증분으로 윈도우 평균템플릿 구성 [그림 3]
Step 3	각각의 장단에 대해 Step 1의 각각의 에너지 국부 최대치 부분을 전후로 Step 2의 각각의 윈도우 평균템플릿에 대해 NCC의 평균치가 가장 큰 윈도우 평균템플릿을 탐색
Step 4	Step 3의 결과로 장단분별 최적의 윈도우 템플릿 생성

[표 2]의 Step 3에서 NCC는

$$NCC(T, F) = \frac{1}{15} \sum_{j=1}^{15} \frac{1}{N} \sum_{i=-N/2+1}^{N/2-1} \frac{(T_j(i) - \bar{T}_j)(F_j(i) - \bar{F}_j)}{\sigma_{T_j} \sigma_{F_j}} \quad (3)$$

와 같이 구할 수 있다[16]. 식 (3)에서  $T$ 는 윈도우 템플릿의 크기  $N$ 을 가지는 대표치 추정 윈도우에서의 15차원 특징벡터 열을 의미하고,  $F$ 는 판소리 음원의 해당위치에서의 특징벡터 열을 의미한다. 또한  $T_j(i)$ 와  $F_j(i)$ 는  $j$ 번째 성분의 윈도우내의  $i$ 위치에서의 특징값을 의미하며  $\bar{T}_j$ 와  $\bar{F}_j$ 는  $j$ 번째 성분들의 템플릿과 음원의 윈도우 내에서의 특징 값의 평균을,  $\sigma_{T_j}$ 와  $\sigma_{F_j}$ 는 표준편차 값을 의미한다. 즉 벡터 열의 NCC 값은 각 성분의 NCC 값의 평균에 해당한다. 또한 [표 2]의 Step 1에서 NLMS(Non-Local Maxima Suppression)란 국부 최대치를 구하는 방법으로 자신보다 크기가 작은 인접 값을 0으로 만드는 연산을 의미한다.

표 3. 최적 윈도우 템플릿 지속시간과 NCC 유사도 값

구분	지속시간	평균(최대)
진양조	1800ms	0.39(0.613)
중모리	1100ms	0.34(0.671)
중중모리	600ms	0.33(0.562)
자진모리	400ms	0.37(0.528)

[표 2]의 산출과정을 통해 얻은 최적의 합(궁) 분별

윈도우 템플릿의 지속시간은 [표 3]과 같다. 즉 표로부터 장단의 빠르기에 따라 합과 궁(진양조의 경우) 위치에서 북의 지속시간이 차이를 보임에 따라 최적의 합(궁) 분별 윈도우의 지속시간이 다르게 나타나는 것을 확인할 수 있다. [표 3]에는 각 장단의 합 위치에서 최적의 윈도우 템플릿에 대한 NCC 값들의 평균과 최대치를 함께 보여주고 있다.

1.4 자동채보를 위한 장단분별

전술한 바와 같이 장단에 따른 최적의 합(궁) 분별 윈도우 템플릿은 모두 다르기 때문에, 이를 이용하게 되면 어떤 종류의 장단인지 찾을 수 있다. 즉 진양조의 합(궁)은 어느 다른 장단과 특성이 다르므로 오로지 최적의 진양조의 분별 윈도우를 이용하여 유사성을 판별하였을 때 높은 유사도를 나타낸다. 따라서 예상되는 여러 합의 위치에서 장단별 윈도우 템플릿과 유사성을 판별하였을 때 유사성이 가장 크고, 소리 부분 전체에서 가장 많이 나타나는 장단이 현재 소리 부분의 장단이라고 판단할 수 있다. 따라서 [표 4]의 알고리즘에서와 같이 소리 부분 전체에서 에너지가 높게 나타나는 부분, 즉 예상 합(궁) 부분을 모두 주러놓고, 그 부분에서 각 장단의 윈도우 템플릿들과 유사성을 판별하여 가장 높은 NCC 값을 제공하는 장단에 투표를 진행하며, 소리 전체에 대해 가장 많이 득표를 한 장단을 다수결의 원칙에 따라 해당 장단으로 판단하였다.

표 4. 다수결에 따른 장단분별 알고리즘

Step 1	장단을 모르는 소리부분에 대해 에너지의 국부 최대치 부분을 NLMS에 의해 탐색함.
Step 2	각각의 에너지 국부 최대치 부분에서 장단별 최적 윈도우 템플릿으로 NCC 유사도 측정 최대의 유사도 값을 가지는 장단에 투표
Step 3	Step 2의 투표결과에 따라 최대득표 장단으로 분별

1.5 자동채보를 위한 마디 찾기

판소리에서 마디개념은 존재하지 않으나 장단에 따라 음길이를 예측하고 음길이를 부호화하기 위해 필요하다. 진양조의 경우 24박이 합 또는 궁의 위치에 따라 4 등분으로 나누어질 수 있으며 6박씩을 한 마디로 가

정하고 3분박이라는 가정 하에 이 마디 길이를 6으로 나눈 길이를 8분 음표 3개로 가정할 수 있다. 마찬가지로 다른 장단에서도 마디라고 간주할 수 있는 부분을 추정할 수 있으며, 합과 다음 합 사이를 마디라고 간주하고 각 마디의 박자수와 2분박 또는 3분박에 따라 음표의 길이를 산출해 낼 수 있다.

[표 5]는 본 논문에서 제안된 마디 찾는 알고리즘을 보여주고 있다. 알고리즘은 크게 두 부분으로 나누어 볼 수 있는데 그 하나는 마디 후보구간을 산출하는 과정(Step 1 - Step 3)과 다른 하나는 칼만 필터에 의한 마디길이의 추정과 탐색과정(Step 4-Step 5)이다.

표 5. 마디 찾기 알고리즘

Step 1	소리부분에 대해 에너지가 문턱치 이상인 부분들에 대해 해당 장단의 최적 윈도우 템플릿과 NCC 유사도 계산
Step 2	NLMS에 의해 Step 1의 NCC 유사도의 국부 최대부분을 마디경계 후보로 탐색
Step 3	Step 2의 경계를 이용하여 연속된 마디 후보구간(마디 경계 및 지속시간) 산출
Step 4	칼만 필터 순방향 동작 연속된 마디길이(마디 경계)의 추정
Step 5	칼만 필터 역방향 동작 마디길이 추정 후 마디구간 보정 또는 삽입

1.5.1 마디후보구간 산출과정

마디 찾기 과정에서는 장단 정보를 활용한다. 따라서 소리의 장단을 분별해내고 실제 마디 찾기의 기준이 되는 합(궁)의 위치를 정확히 추정하는 것이 중요하다. 하지만 판소리에서는 같은 장단이라도 분위기에 따라 북을 부드럽게 치는 경우도 있고 강하고 치는 경우도 있기 때문에 실제 합(궁)위치에서 해당 장단의 최적 윈도우 템플릿과 NCC 유사도 값(Step 1)이 최대가 되지 않는 경우도 많다. 또한 합(궁) 주변에 최적 윈도우 템플릿과 높은 유사성을 가지는 지점이 많이 나타나므로 일단 NLMS 과정을 통해 합(궁)의 마디 경계후보 지점들을 탐색한다(Step 2). 여기서 NLMS 연산결과를 합(궁) 위치 부근의 국부 최대치는 보존하고, 나머지는 전부 삭제한다. 이렇게 되면 합(궁) 위치 주변에 분포한 합(궁) 예상 지점의 수도 줄고 유사성이 없는 합(궁) 예상 지점도 많이 제거된다. [그림 4]는 wave 파일로부터

[표 5]의 Step 1과 Step 2의 과정을 예로서 보여주고 있다. [그림 4]의 마지막 NLMS 연산결과 그림을 보면, NCC 유사도를 보여주는 중간 그림에 반하여 예상지점이 간결화 된 모습을 볼 수 있다. 이렇게 추출된 국부 최대치 부분들은 마디경계 후보가 되며 이들 사이의 간격은 마디지속시간의 후보가 될 수 있다(Step 3).

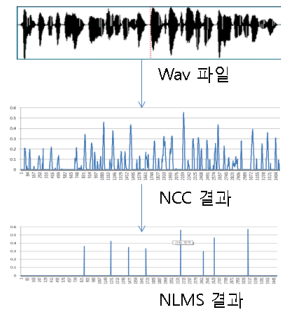


그림 4. NCC와 NLMS를 이용 합(궁) 후보지점 찾기 예시

1.5.2 칼만 필터를 이용한 마디 찾기 알고리즘

관소리의 소리 부분은 중간에 변박이 없기 때문에 마디 구분을 위한 합(궁)의 출현 시기도 일정하게 반복되어진다. 그러나 마디후보구간 산출결과인 마디 후보구간은 템포의 변화를 가정하더라도 매우 불규칙적이다. 왜냐하면 유사도를 바탕으로 한 마디경계와 지속시간의 측정치에 오차를 가지고 있기 때문이다. 뿐만 아니라 경우에 따라서는 고수가 합(궁)을 생략하는 수도 있어 마디의 경계가 존재하지 않아 마디 지속시간이 두 배가 되는 경우도 있다. 따라서 본 논문에서는 합(궁)의 주기 또는 마디 지속시간이 가우시언 분포를 가진다고 가정하고 칼만 필터를 이용하여 마디 지속시간과 마디 경계위치를 추정하였다. 제안된 알고리즘에서 칼만 필터는 정확한 시간 간격을 알지 못해도 시간에 따라 변화하는 추세를 반영하면서 합(궁) 예상 지점의 평균적인 시간 간격을 추정해 준다.

본 논문에서 사용한 선형 칼만 필터는 예측단계와 갱신단계에서 다음 [표 6]과 같이 기술할 수 있다[17].

식 (4)와 (5)에서 아래 첨자 (0, t, t-1 등)는 몇 번째 데이터인가를 의미하며, 윗 첨자 ‘^’은 추정 값을 의미한다. y는 측정 값, x는 상태 값들로 인접한 합(궁)과 합

(궁)사이의 시간간격 즉 마디 지속시간을 의미한다. 식 (5)로 계산되는 갱신된 상태는 1차 저주파 통과 필터(가중 이동 평균) 예측 값과 현재 측정값을 사용하며 제안된 방법의 합(궁) 위치추정의 기반이 된다. F와 H는 시스템 모델 행렬이고 제안된 방법에서는 1 x 1 으로 1을 취하였다. Q와 R은 이들 상태와 출력의 잡음 공분산을 정의하는 잡음 행렬로서 역시 1 x 1으로 Q는 0.0001, 측정 잡음 R은 0.01로 지정하였다. 초기 오차 공분산 P<sub>0</sub>는 1000으로 하였으며, x<sub>0</sub>은 진양조는 11초, 중모리는 10초, 중중모리는 6초, 자진모리는 3초의 장단별 시간 간격을 예측 값 x의 초기 값으로 지정하였다. 해당 초기 값들은 훈련 집합의 음원의 소리부분의 평균적인 장단별 합(궁)의 시간 간격이다.

표 6. 칼만 필터 알고리즘

초기치 선정
$\hat{x}_0, P_0$
예측 단계
$\begin{aligned} \hat{x}_{t t-1} &= F_t \hat{x}_{t-1 t-1} \\ P_{t t-1} &= F_t P_{t-1 t-1} F_t^T + Q_t \end{aligned} \quad (4)$
갱신 단계
$\begin{aligned} \hat{x}_{t t} &= \hat{x}_{t t-1} + K_t (y_t - H_t \hat{x}_{t t-1}) \\ K_t &= P_{t t-1} H_t^T (H_t P_{t t-1} H_t^T + R_t)^{-1} \\ P_{t t} &= (I - K_t H_t) P_{t t-1} \end{aligned} \quad (5)$

제안된 방법에서 최초의 합(궁) 위치 시작은 소리 부분에서의 첫 15초 안(적어도 첫 번째 합(궁)이 예상되는 시기에 NLMS로 추려진 NCC 값이 0.4 이상(안정하게 합(궁)에서의 매칭 값이라고 판단)인 지점을 기준점으로 삼았다. 만약 구간 내에 해당 기준점이 없다면 구간 내에서 가장 큰 NCC 유사도 측정값의 지점을 기준점으로 삼는다. 기준점 이후의 NLMS로 추려진 NCC 값들의 국부 최대 위치들의 시간 간격(즉 예상 마디지속 시간)을 계산하며, 이들을 측정 값 y로 간주하였다(Step 4). 칼만 필터에 의한 추정치들은 정상상태에 도달하기까지는 데이터 변화폭이 크기 때문에 기준점 위치로부터 일정 기간 동안에는 안정된 추정 위치(또는

지속시간)를 제공하지 못한다. 따라서 제안된 방법에서는 소리부분을 뒤에서부터 앞으로 재추정하는 방법을 사용하여 소리의 시작부분의 부정확한 갱신 치들의 정확성을 향상시켰다. 즉 시간이 진행되는 방향(순방향)으로 추정해 나가며 소리의 끝 부분에서 마지막으로 추정된 합(궁)위치를 기준으로 역방향으로 칼만 필터를 다시 적용하였다. 이때 이전 순방향 단계의 칼만 필터 변수 값들을 그대로 계승하여 진행한다.

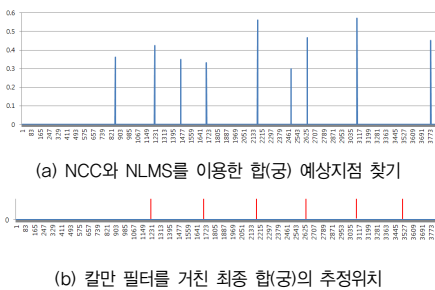


그림 5. 칼만 필터로 보정된 합(궁) 추정 위치

최종적인 합(궁) 위치의 판단은 역으로 진행되는 칼만 필터를 적용할 시(Step 5) 추정된 합(궁)의 지점의 전후 1초 내에 NLMS로 추려진 NCC의 국부 최대치가 존재한다면 해당 지점을 합(궁)의 지점, 즉 마디 분별위치라 판별한다. 그러나 그런 국부 최대치가 존재하지 않는다면 고수가 합(궁)의 복소리를 생략했다고 판단하여 예측된 지점을 마디라 판별하며 시간축의 역방향으로 진행한다. 이러한 역방향 과정을 포함하기 때문에 제안된 알고리즘은 온라인 알고리즘이 될 수는 없으며 오프라인 알고리즘이다. [그림 5]의 (a)는 잘못된 합(궁)의 예상 위치와 검지하지 못했던 합(궁)의 위치를 [그림 5]의 (b)는 칼만 필터를 통해 보정한 결과를 보여주고 있다.

#### IV. 실험 및 검토

본 절에서는 논문에서 제안된 판소리 자동채보를 위한 구조분석 알고리즘들의 성능을 평가하고 검토한다.

#### 1. 사용 판소리 음악

실험에서 사용된 음악은 총 36곡으로 은희진 명창이 부른 춘향가 전집 중 2개의 CD음원으로부터 선택하였다. 선택된 음악은 각각 18개씩 훈련 집합 음악과 검증 집합 음악으로 나누었으며, 16000Hz로 샘플링하고 모노타입으로 변환된 디지털 음성 데이터 파일을 사용하였다. 실험에 사용된 36개의 음원의 총 지속시간은 1시간 16분 25초로 가장 긴 음원이 5분 23초 가장 짧은 음원은 23초였다.

#### 2. 실험 결과 및 토의

실험은 소리/아니리 구분, 장단 분별, 마디분별 등의 항목에 대하여 제안된 알고리즘을 평가하는 방법으로 진행하였다.

##### 2.1 소리와 아니리 구분의 정확도

실험에 사용된 CD음원은 한 트랙 안에 보통 2-3개의 아니리와 소리를 포함하였다. 실험에서는 CD음원의 8개의 트랙을 직접 귀로 들어 소리가 시작되는 시점과 아니리가 시작되는 시점을 체크하여 참값을 산정하였으며, 논문에서 제안한 [표 1]의 알고리즘을 적용하여 예측한 지점이 정확인지 판단하였다. 그 결과 실험에 사용된 음원의 총 26개의 소리와 아니리의 경계 모두를 3초 이내의 정확도(평균  $\pm 1.85$ 초)로 찾을 수 있었다. [표 7]은 탐색된 소리와 아니리 경계의 오차의 분포를 보여주고 있다. 표의 실험결과는 피치 정보가 존재하는 음성구간과 그렇지 않은 비 음성구간의 비율을 활용한 소리부분과 아니리 부분의 구별방법은 적절한 선택임을 알 수 있다. 그러나 비록 3초 이내의 오차라 할지라도 소리부분이 아니리 부분에 포함되면 자동채보 도입부에서 해당부분이 누락되기 때문에 향후 다른 특징들도 함께 고려하여, 오차의 범위를 축소하는 개선방안이 요구된다.

표 7. 소리와 아니리 구분의 시간 오차 분포 수

구분	$\pm(0\sim1)$ 초	$\pm(1\sim2)$ 초	$\pm(2\sim3)$ 초
분포 수	3	11	12
합계	26(개의 지점) (평균 $\pm 1.85$ 초)		



### 2.2 장단분별의 정확도

전술한 바와 같이 [표 3]은 [표 2]에 제안된 방법으로 찾아낸 장단별 합(궁) 위치에서 윈도우 템플릿 백터열의 지속시간과 NCC 유사도 값의 평균치 및 최대치를 보여주고 있다. [표 3]의 결과는 장단의 빠르기에 따라 합(궁) 분별 윈도우 템플릿의 넓이도 달라져야 하며, 고수의 북소리가 지속되는 기간이 느린 장단일수록 길어진다 것을 알 수 있다.

이들 최적의 윈도우 템플릿들을 훈련 집합의 음원들의 예상되는 합(궁)위치에서 유사성을 판단하고 최대의 유사도를 가지는 장단을 찾았을 때 분별의 정확도는 [표 8]의 결과와 같다. 이는 훈련 집합 음원의 소리부분에서 에너지의 국부 최대치 부분, 즉 예상되는 합(궁)의 단일 위치지점을 기준으로 각 장단별 합(궁) 구별 윈도우 템플릿을 개별적으로 적용하여 가장 최대의 유사도를 나타내는 장단을 찾은 결과의 정확도이다.

만약 하나의 예상 합(궁)위치에서 가장 큰 NCC 유사도를 가지는 장단으로 판단하여 정확하게 장단을 맞출 확률이  $p$ 라면  $(1-p)$  만큼은 오차를 포함한다. 또한  $N$ 개의 예상 합(궁) 지점에서  $k$ 번 해당 장단으로 판단될 확률은 이항분포를 따른다. [표 8]과 같이 한 지점에서 장단을 정확히 판단할 확률  $p$ 값이 0.8이상인 경우 다른 장단으로 판단될 확률의 합  $(1-p)$ 는 0.2 미만이며, 한 소리대목에서 해당 장단이 아닌 장단으로 판단되는 횟수는 해당 장단으로 판단되는 횟수보다는 훨씬 작아지기 때문에 논문에서 제안하는 [표 4]의 다수결에 의한 결정은 정확한 분별력을 제공할 수 있다.

표 8. 단일위치 훈련 집합 예상 합(궁) 장단구별 정확도

구분	장단 구별 정확도	예상 합(궁)에서의 장단 구별 적중 수
진양조	88%	112개중 98개
중모리	92%	194개중 178개
중중모리	85%	210개중 184개
자진모리	82%	188개중 154개

[표 2]의 윈도우 템플릿과 [표 4]의 다수결 의사결정 방법에 따른 장단분별 알고리즘을 18개의 검증 집합의 음원에 대해 적용한 결과 100%의 정확도를 보였다. 이는 에너지 값이 국부적으로 최대치를 가지는 합(궁) 예

상위치에서 개별적인 결정의 정확도는 [표 8]의 결과를 제공하였지만 이를 여러 개의 예상 합(궁) 위치에서 투표를 진행하고 다수결로 의사 결정할 경우에는 전술한 이유 때문에 안정된 분별력을 제공한다는 것을 입증한다.

### 2.3 마디구분의 정확도

[그림 6]은 칼만 필터를 이용한 마디위치 추정과정을 보여주고 있다. 그림에서 점선은 연속된 NCC와 NLMS를 이용한 합(궁) 예상지점 간의 시간 간격 즉 마디지속 시간을 나타내고, 순방향으로 칼만 필터를 적용하여 예측한 합(궁) 마디 지속시간은 실선이며, 점선과 가는 실선이 끝나는 부분이 순방향 추정의 끝이다. 그림의 A지점부터 반대방향으로 굵은 선으로 표시된 부분은 역방향 칼만 필터를 적용하여 최종적으로 측정되고(추정 마디 경계의 1초 이내의 측정 경계) 예측(1초 이내의 측정 경계가 존재하지 않을 경우의 예측 경계)된 시간간격이다. 역방향 길이가 순방향의 길이 보다 긴 이유는 순방향의 경우 북소리가 생략된 부분을 역방향의 경우는 예측된 마디 길이로 채웠기 때문이다. 그림에서 북소리가 생략되거나 추정 오차 때문에 측정치를 발견하지 못하고 예측치 자체로 채워진 부분이 약 25%에 해당하는 것을 알 수 있다.

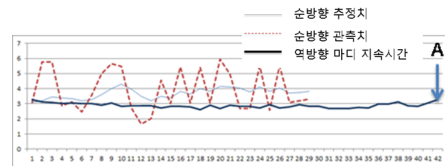


그림 6. 칼만 필터를 이용한 마디 지속시간 예측

[그림 7]은 쑥대머리의 처음 10개의 합(궁)의 시간 간격과 예측된 합(궁)의 시간 간격을 나타낸 것이다. 전술한 바와 같이 중중모리 또는 자진모리 장단은 합(궁) 사이의 간격 즉 마디길이 짧고, 진양조나 중모리 장단은 합(궁) 사이의 간격 즉 마디길이가 길다. 따라서 마디길이 추정 오차를 평균 마디 길이로 나눈 상대오차를 고려하였으며, 제안된 알고리즘의 경우 상대오차가 근사적으로 오차  $\pm 5\%$  범위에서 실험 집합의 모든 음원에서 합(궁), 즉 마디의 정확한 지점을 찾았다.

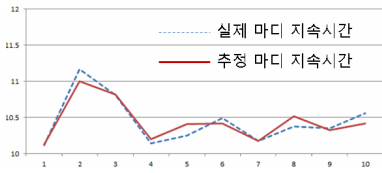


그림 7. 쑥대머리(중모리) 마디 예측 비교

## V. 결론

인터넷을 통한 디지털 음원 서비스가 보편화 되면서 음원 분석과 내용기반 음원검색 연구 역시 큰 흐름을 형성해 가고 있다. 본 논문에서는 유네스코 무형문화유산으로 등재된 판소리의 CD나 녹음된 음원으로부터 자동으로 채보하는 시스템에 필수적인 판소리 구조분석 알고리즘들을 제안하였다. 본 논문에서 제안하는 오프라인 분석기법은 소리와 아니리 부분을 구분하는 알고리즘, 소리부분의 장단 분별알고리즘, 마디를 찾아내는 알고리즘을 포함하였다.

제안된 소리/아니리 구별 알고리즘은 음성과 비 음성 부분의 비율을 특징으로 이용하고 있다. 또한 장단 분별 알고리즘에서는 장단마다 서로 다른 합(궁) 위치에서 북소리 템플릿을 구성하고, 이를 이용하는 다수결 의사결정 방법으로 장단을 강건하게 판단하였다. 뿐만 아니라 마디위치 탐색에서는 예상 마디 경계의 누락과 마디 지속시간의 변화를 극복하기 위해 칼만 필터를 활용하는 방법을 제안하였다.

제안된 방법들은 은희진 명창이 부른 춘향가 판소리 샘플 음원을 이용하여 평가하였으며, 양호한 결과를 제공할 수 있음을 보였다. 제안된 알고리즘을 이용하여 구별된 아니리와 소리, 장단과 마디 경계들은 아니리 부분의 경우 음성인식을 통해 그 내용을 파악하는데 활용할 수 있으며, 소리 부분은 자동채보를 통해 기보하여 유파와 창자의 특징을 분석해 내는데 도움을 줄 것으로 기대된다. 뿐만 아니라 구별된 장단과 더불어 자동채보 후 곡의 가락을 분류한다면 판소리 음원이 담고 있는 정서를 예측하는데도 활용할 수도 있을 것이다. 하지만 향후 우리 소리의 고유의 특징들을 온전히 보존하는 정보화를 위해서는 더 많은 연구가 필요할 것으로

판단된다.

## 참고 문헌

- [1] [http://www.music-ir.org/mirex/wiki/MIREX\\_HOME](http://www.music-ir.org/mirex/wiki/MIREX_HOME)
- [2] Tao Li, Mitsunori Ogihara, and Qi Li, "A comparative study on content-based music genre classification," Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.282-289, Toronto, Canada, 2003.
- [3] C. Xu, N. C. Maddage, and X. Shao, "Automatic music classification and summarization," IEEE Trans. Speech Aud. Processing, Vol.13, No.3, pp.441-450, 2005(5).
- [4] Marolt, "A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music," IEEE Transactions on Multimedia, Vol.6, No.3, pp.439-449, 2004.
- [5] Kenichi Miyamoto, Hirokazu Kameoka, Haruto Takeda, Takuya Nishimoto, and Shigeki Sagayama, "Probabilistic Approach to Automatic Music Transcription from Audio Signals," Proceedings of IEEE Conf. on ASSP, Vol.2, pp.II-697-700, 2007(4).
- [6] 장준영, *퍼지적분을 이용한 곡조 인식 시스템의 설계와 구현*, KAIST 석사 학위 논문, 1996.
- [7] 형아영, 이준환, "SIDE를 이용한 자동 음악채보 시스템", 한국정보처리학회논문지, 제16권, 제2호, 2009.
- [8] 박은중, 신송이, 이준환, "단일 음원 노래에서 음표의 조합 규칙을 이용한 마디 위치 찾기", 한국콘텐츠학회논문지, 제9권, 제10호, pp.1-12, 2009.
- [9] 지정규, *오디오 데이터베이스의 효율적 검색을 위한 선율질의 처리*, 숭실대학교 박사학위 논문, 1998.
- [10] 주영호, 바니야 바부카지, 이준환 "가변 템포를

고려한 자동음악 채보”, 한국콘텐츠학회논문지, 제12권, 제11호, pp.9-19, 2012.

[11] 박형순, *성음 시각기를 이용한 판소리의 채보 연구*, 전북대학교 석사 학위 논문, 2003.

[12] 백대웅, *한국 전통음악의 선율구조 개정판*, 대광문화사, 1982.

[13] MYRON J. ROSS, "Average Magnitude Difference Function Pitch Extractor," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.ASSP-22, No.5, 1974(10).

[14] R. Zhou and J. D. Reiss, "Music onset detection combining energy-based and pitch-based approaches," in *Proc. 8th Int. Conf. Music Information Retrieval*, 2007(9).

[15] B. Logan, "Mel frequency cepstral coefficients for music modeling," In *Proc. Int. Symposium on Music Information Retrieval ISMIR*, 2000.

[16] ETH Zurich, "The Discrete Fourier Transform, Part 6: Cross-Correlation," *Journal of Object Technology*, Vol.9, No.2, 2010(3-4).

[17] <http://www.cs.cornell.edu/Courses/cs4758/2012sp/materials/MI63slides.pdf>

저 자 소 개

주 영 호(Young-Ho Ju)

준회원



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

<관심분야> : 음성처리, 멀티미디어

김 준 철(Joon-Cheol Kim)

정회원



- 1979년 3월 : 전북대학교 전자공학과(공학사)
- 1986년 3월 : 전북대학교 전자공학과(공학석사)
- 1991년 3월 : 전북대학교 전자공학과(공학박사)

▪ 1993년 3월 ~ 현재 : 서남대학교 전자공학과 교수  
<관심분야> : 영상처리, 컴퓨터 비전, 인공지능

서 경 숙(Kyoung-Suk Seo)

정회원



- 1995년 : 전북대학교 예술대학 한국음악과(학사)
- 2000년 : 전북대학교 대학원 음악학과(석사)
- 2010년 : 전북대학교 대학원 한국음악학과 박사과정 수료

▪ 현재 : 전북도립국악원 학예연구사  
<관심분야> : 한국음악, 판소리

이 준 환(Joon-Whoan 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