

인공신경망을 이용한 판소리 마디추정

박근호*, 서경숙**, 이준환*
*전북대학교 컴퓨터공학부
**전북도립국악원
rutepark@naver.com

“Pansori” Joint Assumption using Neural Network

Keunho Park*, Kyoung-suk Seo**, Joonwhoan Lee*

*School of Computer Engineering and Science, Chonbuk National University

**Jeollabuk-do Provincial Institute for Korea Traditional Performing Arts

요 약

본 논문에서는 판소리 자동채보에 중요한 요소인 ‘합’과 ‘궁’의 위치 즉 마디를 인공신경망과 히스토그램을 이용하여 추정한다. 기존의 합과 궁을 추정하는 방법으로는 NCC(Normalized Cross Correlation)를 이용한 대표치 추정 윈도우와 칼만 필터를 이용하였다. 하지만 대표치 추정 윈도우를 구성하는 과정에서 단순히 15개의 특징벡터 각각의 평균을 이용하기 때문에 분별력이 떨어지고, 마디위치를 보정하는 과정에서 칼만 필터를 사용하면 전체음원이 길이가 짧을 경우 오차가 발생할 가능성이 크다. 본 논문에서 제안한 마디 추정 알고리즘은 장단별로 최대 90%이상의 정확도로 마디를 추정할 수 있다.

1. 서론

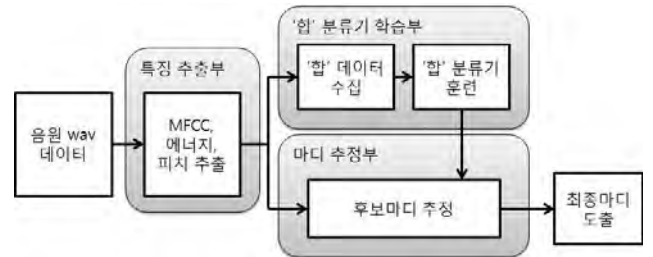
판소리는 2003년 11월 7일 유네스코 무형문화유산으로 등재된 민속음악이며 서민들 사이 유행하던 비정형화된 음악이었기 때문에 체계화되어 있지 못하다.

주영호등의 연구에서는 판소리 음원을 음성과 비음성 구간의 비율로 소리와 아니리로 구분하고, NCC를 이용한 대표치 추정 윈도우를 구성하여 장단분별 하였고, 대표치 추정 윈도우를 이용하여 마디후보구간을 찾은 후, 칼만 필터를 이용 마디후보구간을 보정하여 최종마디를 찾는 방식으로 판소리 자동채보를 진행하였다[1]. 하지만 대표치 추정윈도우를 구성하는 과정에서 단순히 특징벡터 각각의 평균을 이용하기 때문에 분별력이 떨어지고, 최종마디를 찾을 때 사용하는 칼만 필터의 특성상 안정상태에 도달하기까지는 몇 회 정도의 예측과 갱신이 필요하기 때문에 입력된 음원의 길이가 짧은 경우 많은 오차가 발생할 수 있다. 본 논문에서는 제안하는 방법은 오프라인 분석으로, 이미 분류된 ‘소리’를 인공 신경망을 이용하여 장단 구분 없이 후보마디를 찾은 후, 히스토그램을 이용하여 후보마디를 보정하여 최종마디를 찾아내는 알고리즘을 제안한다.

2. 제안된 판소리 마디분석 알고리즘

판소리는 크게 소리와 아니리 부분으로 나누어 볼 수 있다. 아니리 부분은 소리꾼의 독백이므로 채보를 하지 않는다. 본 논문에서는 채보 영역인 진양조, 중모리, 중중모리, 자진모리의 소리 부분을 대상으로 하며, 마디는 ‘합’이라고 하는 북의 가축의 한 가운데인 합구점 타격을 경계로 한다. 본 절에서는 음원에서 이러한 ‘합’의 위치를 분석하여 마디를 찾아내는 과정을 설명하며, ‘합’의 위치에서는

소리꾼의 발성 역시 커지기 때문에 해당 음원의 소리꾼의 발성 정보와 고수의 복소리 정보를 이용한다.



(그림 1) 판소리 자동 마디추정 과정

그림 1은 본 논문에서 제안한 자동 마디분석 과정을 보여 주고 있다. 분석 과정에서는 수집된 판소리를 컴퓨터에서 처리할 수 있는 wav파일(.wav) 형태로 입력하며, ‘합’ 분류기 학습부분과 마디 추정부분으로 나누어 처리한다.

2.1. 특징 추출부

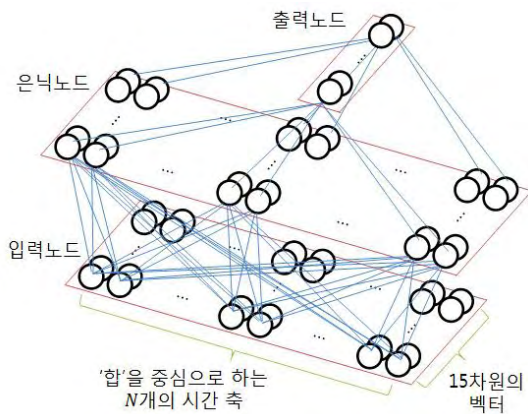
우선 훈련 및 검증 집합의 모든 음원에서 ‘합’과 ‘궁’에 해당하는 복소리와 음성부분을 찾기 위해 피치(pitch), 에너지(Energy), MFCC(Mel-Frequency Cepstral Coefficients) 등의 특징을 추출한다.

피치특징은 음높이를 표현하며 단일음의 주파수와 관계된 지각적인 특성으로 AMDF (Average Magnitude Difference Function)를 이용하여 추출하였다. 본 연구에서 샘플링 주파수는 16000Hz이고 프레임 크기는 320, 창(window) 함수의 넓이는 160으로 설정해 매 10ms 마다 피치정보를 추출하였다[2]. 에너지특징은 시간에 따른 음의 세기(amplitude)를 보여주는 특성으로 일정 크기의 창

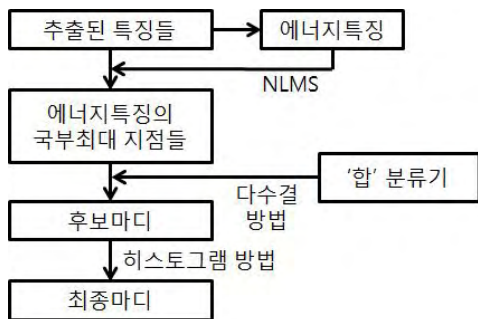
내의 프레임 데이터의 전력(power) 값으로 표현된다[3]. 판소리 구조분석에서 에너지특징은 고수의 북소리의 강약을 검출하거나 소리꾼 발성의 시작부분 또는 강조부분 등을 찾아내는데 유용하다. MFCC특징은 주파수 분포 중에서 인간의 청각시스템에 민감한 부분을 포착하는 13차원의 벡터로 소리분류에 사용되는 대표적인 특징이다[4]. 본 논문의 판소리 구조분석에서 MFCC는 서로 다른 장단에서의 ‘합’ 소리의 특징을 표현하는데 유용하다.

2.2. ‘합’분류기 학습부

제안된 방법에서는 훈련 집합에 속해있는 ‘소리’ 부분에서 전문가가 미리 분석해낸 ‘합’의 위치를 이용하여 학습한다. 전술한 바와 같이 음향의 특징을 나타내는 13차원의 벡터인 MFCC, 에너지, 피치 3가지 특징을 시간 축 상에 늘어놓으면 한 시간 단면에서 15차원의 벡터가 되며, 이를 미리 추정된 ‘합’의 전후에서 매 10ms 단위로 취득하면 그림 2의 한 ‘합’을 중심으로 하는 15차원 벡터 열이 생성된다. 이를 차원별로 각각 따로 훈련하여 15개의 인공신경망 네트워크를 생성한다. 각 네트워크의 입력은 0~1로 정규화된 1개의 시간 축 특징이고, 출력은 값이 클수록 ‘합’에 가까운 정도를 나타내는 0~1사이의 값 1개이다.



(그림 2) ‘합’ 분류기 인공신경망 구조



(그림 3) 마디추정 알고리즘 플로우차트

2.3. 마디 추정부

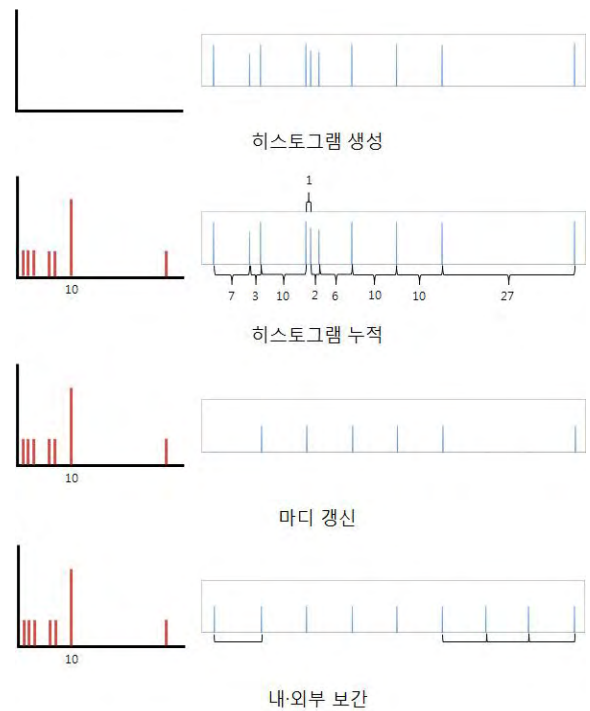
제안된 방법에서는 마디 추정이란 ‘합’과 ‘궁’의 위치추정에 해당하며 그림 3과 같다.

전술한 바와 같이 ‘합’위치에서 고수는 북을 치고, 소리꾼은 발성을 하기 때문에 일반적으로 큰 에너지특징을 가지게 된다. 따라서 추출된 15차 벡터의 특징들 중 에너지특징만을 이용하여 NLMS (Non-Local Maximum Suppression) 방법으로 국부최대(Local Maxima)를 구하고, 국부최대 지점들을 ‘합’후보로 한다.

‘합’후보 지점들을 중심으로 전후에서 N 개만큼의 15차원 벡터를 추출하고 각 차원마다 입력노드에 N 개의 특징을 입력하면 15개의 출력 값을 얻을 수 있다. 다수결의 원칙에 따라 이 출력 값들 중 0.5보다 큰 값들이 8개 이상이면 ‘합’후보로 남겨지나 아닐 경우 ‘합’후보에서 탈락된다. 이 ‘합’후보들의 집합을 후보마디로 한다.

판소리의 마디는 장단마다 같은 길이를 가진다. 하지만 소리꾼은 항상 완벽한 템포로 노래를 하지 않고, 고수는 흥에 겨우면 가끔씩 ‘합’지점에서 북을 타격하지 않기도 한다. 따라서 마디의 시작점인 ‘합’위치는 절대 적이지 못하고 구간에 따라서 상대적이다. 이러한 상대적인 ‘합’위치를 명확하게 찾아내기 위해 본 논문에서는 히스토그램방법을 제안하였으며, 과정은 그림 4와 같다.

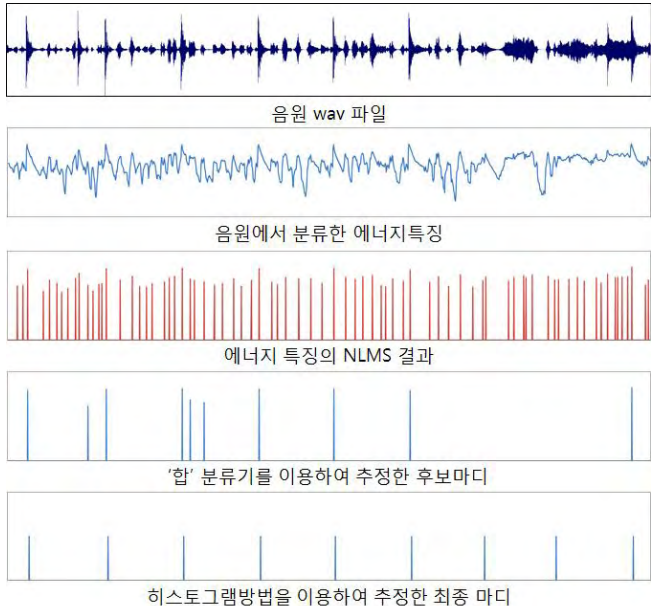
히스토그램방법의 수행과정은 다음과 같다. 각 빈(bin)들이 0으로 초기화된 히스토그램을 생성한다. 후보마디들을 각 길이에 해당하는 히스토그램의 빈에 누적한다. 가장 수치가 높은 빈에 누적된 마디를 제외한 나머지 마디는 삭제한다. 남은 마디를 중심으로 내·외부 보간을 실시하면 마디추정이 완성된다.



(그림 4) 히스토그램을 이용한 마디추정 과정

그림 5는 자진모리장단의 마디추정의 결과를 단계적으로 보여주고 있다. 국부최대치를 이용하여 음원 전역에서

탐색하지 않아도 되어 계산량이 감소하고, 인공신경망을 이용한 ‘합’ 분류기로 합에 가장 가까운 특성을 찾을 수 있어 육안으로 피치와 에너지 정보만 가지고 자칫 합으로 판단할 수 있는 부분에서 오판을 하지 않으며, 히스토그램 방법을 이용하여 고수의 북소리가 생략된 부분도 보간을 하여 마디를 구성하는 것을 볼 수 있다.



(그림 5) 마디추정 과정별 결과들

3. 실험 및 검토

본 절에서는 논문에서 제안된 판소리 마디분석 알고리즘들의 성능을 평가하고 검토한다.

3.1. 사용 판소리 음악

실험에서는 이일주명창이 부른 춘향가 재회편 중 장단별 ‘소리’음원 16개를 선택하였고, 16000Hz로 샘플링하고 모노타입으로 변환된 디지털 음성 데이터 파일을 사용하였다. 실험에 사용된 16곡의 음원의 총 지속시간은 47분 26초로 가장 긴 음원이 8분 58초 가장 짧은 음원은 27초였다.

3.2. ‘합’위치 탐색과 마디구분의 정확도

실험은 두 가지로 나누어 진행하였다. 첫 번째는 인공신경망 ‘합’분류기의 탐색의 정확도를 측정하기 위해 데이터를 훈련집단과 실험집단으로 나누어 수행하였고, 표 1는 첫 번째 실험결과와 혼동행렬(confusion matrix)을 보여주고 있다. ‘합’의 참 위치는 약 85%의 정확도로 예측을 하였으며, ‘합’의 거짓 위치는 약 99%의 정확도로 예측을 하였다. 두 번째는 마디구분의 정확도를 측정하기 위한 실험으로 전문가가 찾아낸 ‘합’의 위치를 본 논문의 알고리즘이 얼마나 정확히 일치 시키는가를 측정하였다. 표 2은 두 번째 실험의 결과이다. 실험 음원에서 진양조에서 전반적

으로 ‘합’ 위치에서의 북소리가 많이 생략되어 저조한 정확도를 보이는 것으로 판단되며 그렇지 않은 나머지 음원들에서는 평균 80%가 넘는 정확도를 보였다.

4. 결론

본 논문은 먼저 연구된 주영호의 논문에서 마디추정을 칼만 필터로 하기 때문에 음원의 길이가 짧을 때 오차가 많이 발생한다는 단점을 히스토그램 방법을 이용하여 보완하였고, 장단 구분 없이 바로 ‘합’의 위치를 인공신경망을 이용하여 분별해내는 새로운 방법을 제안하였다.

제안된 방법들은 이일주명창이 부른 춘향가판소리 샘플 음원을 이용하여 평가하였으며, 양호한 결과를 제공할 수 있음을 보였다. 진양조의 경우 고수가 ‘합’의 위치에서 합궁점 타격을 생략한 경우가 많아서 상대적으로 낮은 정확도 보였지만, 에너지의 국부최대점만 분석하는 것이 아니라 음원의 전반적인 파형을 분석하여 시간에 따른 변화를 평가하여 장단을 예측한다면 마디위치 추정이나 전체적인 판소리 분석에 더 나은 결과를 낼 수 있을 것이다.

<표 1> ‘합’ 위치탐색 실험결과

분류결과 / 참 부류	참	거짓
참	162(85.26%)	28(14.74%)
거짓	2(0.52%)	382(99.48%)

<표 2> 마디 위치추정 실험결과

장단	진양조	중모리	중중모리	자진모리
정확도	65.38%	91.67%	80.00%	82.35%

참고문헌

- [1] 주영호, 김준철, 서경숙, 이준환, “판소리 자동채보를 위한 구조분석 알고리즘”, 한국콘텐츠학회논문지, 제 14권 제2호, pp.28-38, 2014.
- [2] MYRON J. ROSS, “Average MagnitudeDifference Function Pitch Extractor,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-22, No.5, 1974(10).
- [3] 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).
- [4] B. Logan, “Mel frequency cepstral coefficients for music modeling,” In Proc. Int. Symposium on Music Information Retrieval ISMIR, 2000.