논문 2010-47SP-5-13

하모닉 구조를 이용한 다성 음악의 주요 멜로디 검출

(Extracting Predominant Melody from Polyphonic Music using Harmonic Structure)

윤 제 열*, 이 석 필**, 서 경 학**, 박 호 종***

(Jea-Yul Yoon, Seok-Pil Lee, Kyeung-Hak Seo, and Hochong Park)

요 약

본 논문에서는 하모닉 구조를 이용하여 다성 음악의 주요 멜로디를 검출하는 방법을 제안한다. 다성 음악은 다수의 음원을 동시에 포함하므로 주요 멜로디를 검출하기 위하여 다중 기본 주파수를 추출하고 각 기본 주파수의 성질을 기반으로 주요 멜로디를 구하는 과정으로 구성된다. 하모닉 구조는 기본 주파수의 배음관계를 나타내고 단일 음원 신호의 중요한 특성 파라미터이다. 따라서 제안하는 방법은 하모닉 구조의 정확도를 기준으로 다성 음악에 존재하는 모든 기본 주파수 후보를 추출하고, 추출된 기본 주파수 후보에 대하여 하모닉 성분을 조합하여 하모닉 평균 에너지를 구하여 기본 주파수 후보의 중요도 순위를 결정한다. 마지막으로 기본 주파수 후보의 순위와 기본 주파수의 연속성을 기반으로 피치 트래킹을 진행하여 최종 주요 멜로디에 해당하는 기본 주파수를 검출한다. 제안한 방법의 성능을 ADC 2004 DB와 가요 100곡에 대하여 MIREX 2005 측정 방법에 따라 측정하였으며, ADC 2004 DB에 대하여 90.42%의 검출 정확도를 가진다.

Abstract

In this paper, we propose a method for extracting predominant melody of polyphonic music based on harmonic structure. Since polyphonic music contains multiple sound sources, the process of melody detection consists of extraction of multiple fundamental frequencies and determination of predominant melody using those fundamental frequencies. Harmonic structure is an important feature parameter of monophonic signal that has spectral peaks at the integer multiples of its fundamental frequency. We extract all fundamental frequency candidates contained in the polyphonic signal by verifying the required condition of harmonic structure. Then, we combine those harmonic peaks corresponding to each extracted fundamental frequency and assign a rank to each after calculating its harmonic average energy. We finally run pitch tracking based on the rank of extracted fundamental frequency and continuity of fundamental frequency, and determine the predominant melody. We measure the performance of proposed method using ADC 2004 DB and 100 Korean pop songs in terms of MIREX 2005 evaluation metrics, and pitch accuracy of 90.42% is obtained.

Keywords: 다성 음악(polyphonic music), 주요 멜로디 추출(predominant melody extraction), 다중 피치 추출(multi-pitch extraction), 하모닉 구조(harmonic structure)

I. 서 론

음악 데이터를 효율적으로 분석하고 검색하기 위하여 CASA(computational auditory scene analysis), 다중 피치 추출(multi-pitch extraction), QbH(query by humming) 등에 대한 관심이 커지고 있다. 이들을 위하여 다성 음악(polyphonic music)에서 주요 멜로디 혹은 보컬 멜로디를 추출하는 과정이 공통적으로 필요하며, 이를 위한 많은 연구가 진행되고 있다.

음악신호에 대한 정보처리는 MIREX(music information retrieval exchange)를 중심으로 많은 연구

* 학생회원, *** 정회원, 광운대학교 전자공학과
(Kwangwoon University)
** 정회원, 전자부품연구원
(Korean Electronics Technology Institute)
접수일자: 2010년7월5일, 수정완료일: 2010년8월5일

와 기술 교류가 진행되고 있으며 MIREX의 주요 활동은 [1]에 정리되어 있다. 멜로디 추출을 위한 핵심 기술인 다중 피치 검출에 대한 기존 기술은 [2]에 정리되어 있다. 또한, 멜로디를 추출하기 위하여 Goto는 PreFEst (predominant-F0 estimator)를 제안하였으며[3], 예측에 사용되는 MAP(maximum a posteriori) 방법에 EM (expectation maximization) 방법을 적용하여 다중 연결(multiple tracking)을 위하여 시간적 연속성을 고려하는 방법도 제안하였다[4].

Klapuri는 주파수 영역에서 이상적인 하모닉 구조의 성질, 즉 기본 주파수의 정수배에 하모닉 피크가 나타나는 성질을 이용하고, 반복적으로 강한 피치를 구하고 제거하는 방법을 제안하였고, 평탄화(smoothing) 과정을 첨가하여 비조화(inharmonic) 문제점을 해결하였다[5]. 또한 Klapuri는 [5]를 기반으로 향상된 새로운 방법을 제안하였다[6]. 이 방법은 기존의 기술들에 비하여 간단하면서도 효율적이고 하모닉 진폭의 합(summing harmonic amplitudes)을 기반으로 하며, 우수한 성능을 가지는 대표적인 다중 피치 추출 기술이다.

Lagrange는 기존 방법들과는 다르게 텍스처 윈도우즈(texture windows)를 사용하여 스펙트럼 특성을 표현하고 개별 피크의 파라미터인 주파수, 진폭, HWPS(harmonically wrapped peak similarity)를 이용하여 유사도를 연산하여 집단화(cluster) 하는 방법으로 멜로디를 추출한다[7]. 이때 N-cut(normalized cut)을 이용하여 클러스터들의 경계를 나누게 되고, 이렇게 나누어진 클러스터를 기존에 저장된 음색 모델들과 매칭을 통하여 클러스터의 음원을 구분하고 보컬 멜로디를 추출한다. Zhang은 하모닉 구조의 AHS(average harmonic structure)와 HSS(harmonic structure stability)를 추출하고 각각의 하모닉 구조를 집단화하는 방법을 사용하고, Lagrange와 마찬가지로 단성 형태로 개별 악기의 음색을 모델링하여 정합하는 방법으로 음원을 구분한다[8].

Durrieu은 STFT(short time Fourier transform)를 사용하여 TFR(time-frequency representation)을 만들고 이 신호를 GMM(Gaussian mixture model) 적용하여 보컬 모델링과 배경음 모델링을 하며, 다성 음악은 보컬과 배경음으로 나누어진다는 전제하에 두 가지로 모델 분석한다[9]. Vincent는 NMF(non-negative matrix factorization) 방법을 적용하여 스펙트럼에서 NMF 모델을 사용하는 방법을 제안하였다[10]. NMF는 입력 신호를 동조(tuning) 모델에 따라 하모닉 집단과 비조화 집단으로 모델링한다. 3개의 동조 모델과 2개의 배음(overtone) 모델을 결합하여 총 6개의 NMF 모델을 만들고, NMF의 특성에 따라 하모닉과 비조화 모델링된다. 입력 신호의 특성에 따라 NMF의 파라미터는 적응적으로 반응한다.

이상의 기존 기술들은 우수한 성능을 보이지만 다음과 같은 문제점이 존재한다. 통계적 성질을 분석하여 HMM(hidden Markov model), GMM(Gaussian mixture model)을 적용하여 확률 모델을 이용한 방법은 일반적으로 신호 단독의 데이터를 이용하여 모델링하여 모델 신호 추출에 높은 성능을 보이지만 모델 신호와 이질적인 신호를 포함한 합성 신호에 대하여 낮은 성능을 보이며, 모델을 구하기 위하여 매우 복잡한 훈련과 검색 과정이 필요하여 매우 많은 연산량이 요구된다. 기존의 하모닉 구조를 이용한 방법은 안정적인 하모닉 구조에 대하여 뛰어난 성능을 보이지만 많은 악기의 합성으로 하모닉 구조가 정확한 정수배를 이루지 않고, 부분적으로 하모닉 구조가 손실될 경우 성능이 저하되며, 찾은 기본 주파수의 하모닉 성분을 감쇄 혹은 제거하여 새로운 기본 주파수를 검색하는 반복 동작이 진행됨에 따라 스펙트럼 정보가 점차 제거되어 정확한 기본 주파수 추출에 어려움이 있다.

본 논문에서는 기존 기술들의 문제점 및 한계를 해결하기 위하여 신호의 복잡한 모델링 과정 없이 간단한 방법으로 기본 주파수를 추출하며, 하모닉 구조를 기반으로 모든 피크 위치를 사용하여 기본 주파수를 추출하고 기본 주파수의 연속성을 이용한 주요 기본 주파수 추출 방법을 제안한다. 특히, 가장 대표적인 내용 기반 음악 검색 시스템인 QbH 시스템에 적합한 멜로디 검출 방법을 제안한다. 일반적으로 가요에서는 보컬 영역뿐만 아니라 반주 구간에서도 주요 멜로디가 존재하는 경우가 많으며, 사람은 보컬의 유무에 관계없이 노래에서 반복적인 대표 멜로디를 주로 기억한다. 따라서 본 논문에서는 간단하고 효과적인 방법으로 보컬과 보컬이 없는 구간의 구분 없이 모든 구간에서 주요 멜로디를 검출하는 방법을 개발한다. Ⅱ장에서는 제안하는 주요 멜로디 검출 방법의 다중 피치 추출과 피치 트래킹 방법에 대하여 설명하고, Ⅲ장에서는 제안한 방법의 성능 측정 결과를 보여준다. 마지막으로 Ⅳ장에서는 결론에 대해서 논의한다.

Ⅱ. 주요 멜로디 추출 방법

다성 음악은 다수의 음원(source)을 동시에 포함하므로 주요 멜로디를 검출하기 위하여 먼저 다중 피치 주파수를 추출하고, 추출된 다중 피치 주파수 중에서 주요 멜로디에 해당하는 기본 주파수를 선택하는 과정이 필요하다. 이러한 동작을 효과적으로 구현하기 위하여 본 논문에서는 음악 신호의 중요한 특성인 하모닉 구조를 기반으로 다중 피치 추출과 주요 멜로디 추출 방법을 제안한다. 그림 1은 본 논문에서 제안하는 주요 멜로디 검출 방법의 전체 구조를 나타낸다.

음악 신호의 기본 주파수는 낮은 주파수 대역에서 결정되므로 고대역 성분은 기본 주파수 검출에 영향을 미치지 않는다. 따라서 전처리 모듈에서는 입력된 음악 신호를 8kHz 샘플링 주파수로 다운 샘플링(down sampling) 한다. 또한, 제안하는 방법을 모노와 스테레오에 관계없이 범용으로 사용하기 위하여 스테레오 입력일 경우 신호를 모노로 변환한다. 이렇게 전처리된 신호를 50% 중첩을 가지는 16ms 단위의 프레임으로 구분하여 Hanning 윈도우를 적용하고 N-포인트 DFT(discrete Fourier transform) 하여 스펙트럼 $X[k]$을 구하고 다음 동작을 수행한다.

다중 피치 추출 모듈은 신호에 포함되어 있는 다수의 피치 정보를 추출하고 추출된 피치들의 하모닉 구조의 유무와 정확도에 따라 유효한 피치를 선택하고 각각의 순위를 정한다. 먼저 신호에 포함된 가능한 모든 의미 있는 기본 주파수 후보들을 찾는다. 하모닉 피크일 확률이 높은 주파수 피크를 찾고 피크들 사이의 간격 분포를 분석하여 해당 피크가 하모닉 구조의 조건을 만족하는지 확인하고, 조건을 만족하는 기본 주파수 후보를 찾는다. 다음, 하모닉 구조 집단화 모듈은 추출된 모든 기본 주파수 후보에 대하여 각각의 하모닉 성분을 결합하여 하모닉 그룹을 생성한다. 마지막으로 각 하모닉 그룹의 평균 에너지를 구하여 각 기본 주파수의 중요도 순위를 결정한다.

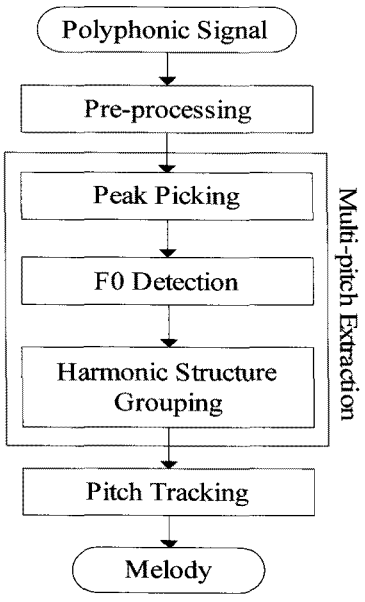

그림 1. 제안하는 방법의 전체 구조
Fig. 1. Overall Structure of the proposed method.

피치 트래킹 모듈은 이전 및 이후 프레임의 기본 주파수의 연속성과 기본 주파수의 순위를 고려하여 피치 트래킹을 실행하고, 최종 주요 멜로디에 해당하는 기본 주파수를 선택한다. 본 논문에서 제안하는 주요 멜로디 검출 방법의 각 모듈에 대한 자세한 설명은 다음과 같다.

1. 하모닉 기반 다중 피치 추출 방법

가. 피크 피킹(peak picking)

다성 음악 신호는 다수의 음원들이 혼합되어 있는 신호이며, 각 음원의 기본 주파수에 해당하는 하모닉 피크들의 조합이 존재한다. 따라서 기본 주파수를 추출하기 위하여 먼저 스펙트럼에 나타나는 주파수 피크를 검출한다. 본 논문에서는 식 (1)의 조건을 사용하여 $X[k]$의 실제 하모닉 피크에 해당할 확률이 높은 주파수 피크를 검출한다.

$$\begin{aligned} &|X[k]| > |X[k-1]| \text{ and} \\ &|X[k]| > |X[k+1]| \text{ and} \\ &|X[k]| > PTH_{(l \text{ or } h)} \end{aligned} \tag{1}$$

여기서 PTH는 피크 선택의 경계값으로 피크 검출의 성능을 결정하는 중요한 파라미터이며, 잘못된 피크를 검출하지 않으면서 주요 피크를 누락하지 않도록 한다. PTH_l, PTH_h는 각각 저대역과 고대역의 피크 경계값을 의미하고 저대역과 고대역은 2kHz를 기준으로 나눈다. 일반적인 음악 신호는 저대역과 고대역의 평균적인 에너지에 차이를 가진다. 따라서 해당 프레임의 주파수 Skewness(SK)를 구하여 주파수 포락선 모양에 따라 PTH를 가변적으로 정한다. SK는 식 (2)와 같이 정의되며, $\overline{X}$는 $|X[k]|$의 평균값이다.

$$SK = \sum_{k=0}^{N/2} (|X[k]| - \overline{X})^3 \tag{2}$$

$SK=0$이면 대칭적(symmetric) 분포, $SK<0$이면 고대역에 더 많은 에너지 분포, $SK>0$이면 저대역에 더 많은 에너지 분포를 의미한다. 따라서 $SK=0$ 이면 $PTH_l, PTH_h = \overline{X_a}$이고, $SK<0$ 이면 $PTH_l = \overline{X_a} - \sigma_a$, $PTH_h = \overline{X_h} - \sigma_h/2$, $SK>0$ 이면 $PTH_l = \overline{X_a} - \sigma_a/2$, $PTH_h = \overline{X_h} - \sigma_h$ 를 사용한다. 여기서 $\overline{X_a}$, $\overline{X_h}$는 각각 전대역과 고대역의 평균값이고, σ_a, σ_h는 각각 저대역과 고대역의 표준 편차값을 의미한다.

나. 기본 주파수 추출

앞에서 검출한 주파수 피크의 위치를 $localpeak[\bullet]$라 할 때, 기본 주파수 후보를 정하기 위하여 식 (3)과 같이 정의된 피크 사이의 간격을 사용한다.

$$\Delta[i,j] = localpeak[i] - localpeak[j] \quad (3)$$

여기서 $i = j+1,\ldots,C$, $j = 1,\ldots,C$ 이고, C 는 해당 프레임에서 추출된 주파수 피크의 전체 개수를 나타낸다. $\Delta[i,j]$은 150Hz~1kHz 영역으로 제한한다. 다음, $\Delta[i,j]$가 기본 주파수의 조건을 만족하는지를 하모닉 구조를 기반으로 확인한다. 즉, 모든 기본 주파수는 이상적인 하모닉 구조를 가진다는 가정하고 각 기본 주파수 후보에 대하여 해당 하모닉 피크 위치에 실제 주파수 피크가 존재하는지 확인한다. 이상적인 하모닉 피크 위치 $h_p[\bullet]$는 식 (4)에 따라 정의한다.

$$h_p[m] = localpeak[j] + (m-d) \times \Delta[i,j] \quad (4)$$

이 때, $m = 0,\ldots, N/(\Delta[i,j] \times 2)$이고, d는 하모닉 피크의 시작점을 결정하는 파라미터로서 식 (5)와 같이 정의된다.

$$d = INT(localpeak[j]/\Delta[i,j]) \quad (5)$$

이렇게 구한 이상적 하모닉 피크 위치 $h_p[\bullet]$에 대하여 $h_p[\bullet] \pm B$ 영역에 실제 주파수 피크 $localpeak[\bullet]$가 존재하는지 확인하고, 이에 따라 $h_p[\bullet]$에서의 조건 만족 여부를 정한다. 본 논문에서는 B = 15Hz를 사용한다. 이와 같은 방법으로 모든 $\Delta[i,j]$ 각각에 대하여 모든 이상적 하모닉 피크 위치 $h_p[\bullet]$에 대한 조건 만족 여부를 확인하고 50%이상의 하모닉 피크에서 조건이 만족되면 해당 $\Delta[i,j]$를 기본 주파수 후보 $\widehat{F0}$ 로 결정한다. 만일 프레임 내에 위 조건을 만족하는 $\widehat{F0}$ 가 없으면 해당 프레임은 멜로디가 없는 프레임으로 판별한다.

다. 하모닉 구조 집단화

앞에서 설명한 다중 피치 추출 방법으로 추출된 기본 주파수 후보 $\widehat{F0}$에 대하여 해당 하모닉 구성원을 스펙트럼 $X[k]$에서 다시 선택한다. 먼저, $\widehat{F0}$의 이상적 하모닉 피크 위치에 대하여 $h_p[\bullet] \pm B$ 영역에서 가장 큰 스펙트럼 크기 $|X(k)|$ 를 $\widehat{F0}$의 하모닉 구성원으로 정한다. 이 때, 두 개의 기본 주파수가 정수배 관계를 가지면 두 개 모두 기본 주파수로 선택될 가능성이 있으므로 피치 더블링(pitch doubling)이 발생한다. 이를 방지하기 위하여 서로 다른 $\widehat{F0}$에 대하여 하모닉 피크 위치가 중복되는 형태를 분석하고 하모닉 구성원이 85%이상 같은 경우에는 같은 음원으로 판단하고 큰 $\widehat{F0}$는 기본 주파수 후보에서 제외한다.

프레임 단위로 구해진 기본 주파수 후보의 중요도 순위를 결정하기 위하여 AHS(average harmonic structure) 값을 계산한다[8]. 즉, 하모닉 집단화 과정에서 구한 각 $\widehat{F0}$의 하모닉 피크값의 평균 에너지를 구하고, 이 값의 크기 순서에 따라 $\widehat{F0}$의 중요도 순위를 정한다.

라. 다중 피치 추출 방법의 성능 측정

본 논문에서 제안하는 다중 피치 추출 방법에 대한 성능을 측정하였다. 각 프레임에 대하여 정답 기본 주파수가 다중 피치 추출을 통하여 1순위로 선택되는 것이 목표이므로 TREC Q&A(text retrieval conference question answering) 측정에 사용되는 MRR(mean reciprocal rank)를 적용하여 다중 피치 추출의 성능을 측정하였다[11]. MRR은 식 (6)과 같이 정의된다.

$$MRR = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{rank_n} \quad (6)$$

여기서 N은 프레임의 전체 개수를 의미하고, $rank_n$은 n번째 프레임의 정답 기본 주파수에 해당하는 $\widehat{F0}$의 순위를 나타낸다. 정답 기본 주파수의 1/4 톤(tone)을 기준으로 평가하였고, ADC(audio description contest) 2004 DB를 사용하였다.

그림 2는 ADC 2004 DB 중 pop1~pop4의 MRR 측정 결과를 보여주고, MRR의 평균값은 0.86이다.

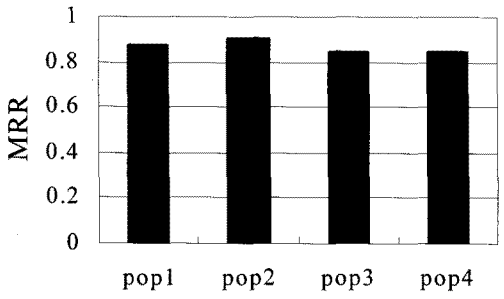

그림 2. MRR 측정 결과
Fig. 2. MRR results.

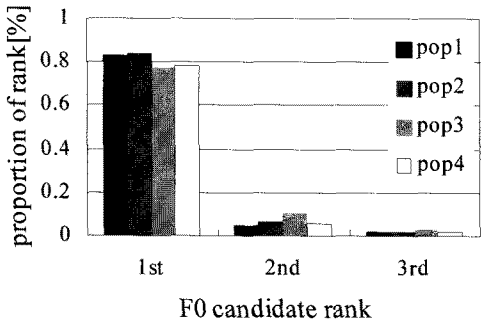

그림 3. 기본 주파수 순위의 히스토그램
Fig. 3. Rank histogram of fundamental frequency.

그림 3은 기본 주파수 순위의 히스토그램을 나타낸다. 제안하는 다중 피치 추출 방법으로 사용하면 75% 이상의 프레임에서 정답 기본 주파수가 1순위로 검출되는 것을 보여준다.

2. 피치 트래킹

각 프레임 별로 추출한 기본 주파수 후보의 피치 트래킹을 통하여 주요 기본 주파수를 선택한다. 피치 트래킹 모듈은 주변 프레임간의 기본 주파수의 연속성을 기반으로 동작하며 처리 과정은 다음과 같다.

- 현재 프레임의 1순위 기본 주파수를 기준으로 이전 프레임과 이후 프레임과의 기본 주파수 연속성을 측정한다.
- 현재 프레임의 1순위 기본 주파수가 연속적이지 않을 경우, 이전 프레임과 이후 프레임의 기본 주파수가 같으면 현재 프레임의 기본 주파수는 이전 프레임의 기본 주파수로 한다.
- 현재 프레임의 1순위 기본 주파수가 연속적이지 않고, 이전 프레임과 이후 프레임의 기본 주파수가 같지 않으면 현재 프레임의 두 번째, 세 번째 순위 기본 주파수와의 연속성을 측정하여 연속적인 기본 주파수를 현재 프레임의 기본 주파수로 한다.
- 위에서 처리되지 않은 경우는 1순위 기본 주파수를 사용하여 새로운 음의 시작점으로 결정한다.

III. 성능평가 및 고찰

성능 평가를 위하여 ADC 2004 DB의 pop을 사용하였다. ADC 2004 DB는 총 20개의 모노 데이터(44.1kHz 샘플링 주파수, 16bit PCM)로 구성되어 있다. 정답으로 사용되는 멜로디의 기본 주파수는 SMSTools를 사용하여 5.8ms 단위로 추출되어 제공된다. 본 논문에서 제안하는 멜로디 추출 방법은 16ms 단위로 기본 주파수가 추출되므로 정답 기본 주파수를 0.2ms 단위로 나눈 후, 정답 기본 주파수 80개에 대한 평균 주파수 값을 성능 측정에 사용하였다.

성능 측정 방법은 [12]에 제시된 MIREX 2005 측정 방법에 따라 진행 하였고, 멜로디 검출 성능은 RPA(raw pitch accuracy)와 RCA(raw chroma accuracy)로 측정하였다. RPA는 정답 멜로디 주파수에 대한 측정된 멜로디 주파수의 정확도를 나타내고 식 (7)과 같이 정의된다.

$$\mathrm{RPA} = \frac{\mathrm{TPC} + \mathrm{FNC}}{\mathrm{GV}} \tag{7}$$

여기서 GV는 보컬 멜로디가 존재하는 프레임의 개수를 나타낸다. TPC는 GV 개 프레임 중에서 주파수를 정확하게 추출한 프레임의 개수를 의미하고, 정답 멜로디 주파수의 ±1/4 톤까지 정확한 멜로디 주파수를 추출한 것으로 결정한다. FNC는 보컬 구간이 아니라고 판정하였지만 정답 멜로디 주파수를 정확히 추출한 프레임의 개수를 나타낸다. 본 논문에서는 보컬 구간을 별도로 검출하지 않으므로 항상 FNC = 0이다. RCA는 추출한 주파수의 옥타브 오류를 무시하고 RPA와 같은 방법으로 주파수 정확도를 측정한 것이다.

표 1은 MIREX 2009의 멜로디 추출 분야 참가자들의 성능과 제안한 방법의 성능을 보여준다. 본 논문에서 제안하는 방법의 성능이 모든 참가자들 보다 우수한 것을 알 수 있다. MIREX 2009의 참가자의 경우 보컬 구간과 보컬이 없는 구간을 나누어 보컬의 경우에만 멜로디를 나타내었다. 그러나 본 논문에서는 모든 구간에서

표 1. MIREX 2009 오디오 멜로디 측정 결과 -ADC 2004 DB(보컬). (from[13])

Table 1. Results of MIREX 2009 audio melody extraction. -ADC 2004 DB(vocal). (from[13])

Participant	RPA(%)	RCA(%)
Cao and Li	85.625	86.205
Durrieu & Richard	86.960	87.398
Hsu, Jang & Chen	63.110	74.101
Joo, Jo & Yoo	81.959	85.798
Dressler	85.969	86.424
Wendelboe	83.135	86.593
Cancela	86.962	87.545
Rao and Rao	81.446	88.038
Tachibana, Ono, Ono & Sagayama	59.768	72.129
Proposed Method	90.418	92.27

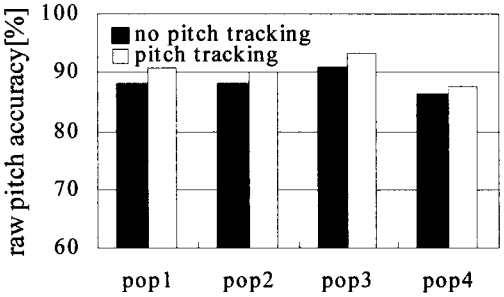

그림 4. 피치 트래킹에 따른 성능 향상

Fig. 4. Performance enhancement by pitch tracking.

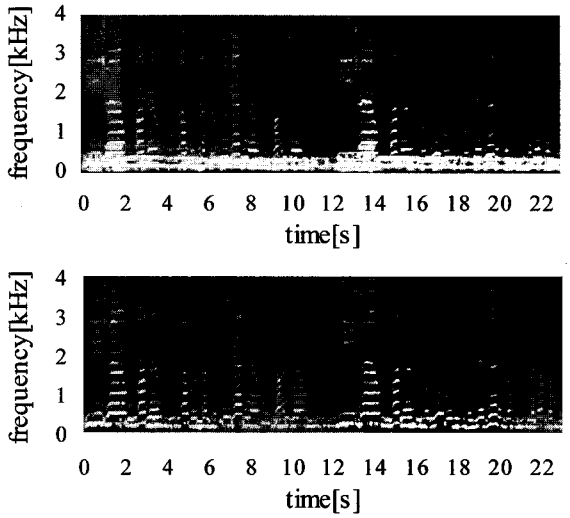

그림 5. ADC 2004 DB (pop1)에 대한 멜로디 검출 결과 (상단: 원 신호의 스펙트로그램, 하단: 제안하는 방법으로 검출된 멜로디 성분의 스펙트로그램)

Fig. 5. Melody extraction result of ADC 2004 DB (pop1). (upper: original spectrogram, lower: spectrogram of extracted melody by the proposed method).

주요 멜로디를 추출하기 때문에 좀 더 높은 성능을 보이는 것으로 판단된다.

그림 4는 피치 트래킹의 적용에 따른 성능 향상을 구체적으로 보여준다. 피치 트래킹을 통하여 1.3~2.8% 정확도가 향상되었음을 알 수 있다.

그림 5는 ADC 2004 DB (pop1)에 대하여 본 논문에서 제안하는 멜로디 추출 결과를 스펙트로그램으로 보여 준다. 밝은 부분은 높은 에너지를 의미하고 어두운 부분은 낮은 에너지를 나타낸다. 여기서, 프레임별로 앞에서 구한 멜로디 주파수의 하모닉 성분들만을 IDFT하여 멜로디 신호를 합성하고, 이 신호의 스펙트로그램을 구하였다. 0~2s, 12~14s 구간은 복잡도가 심하지만 본 논문에서 제안하는 방법을 통하여 효과적으로 멜로디 주파수를 검출하였음을 알 수 있다. 일반적으로 0~

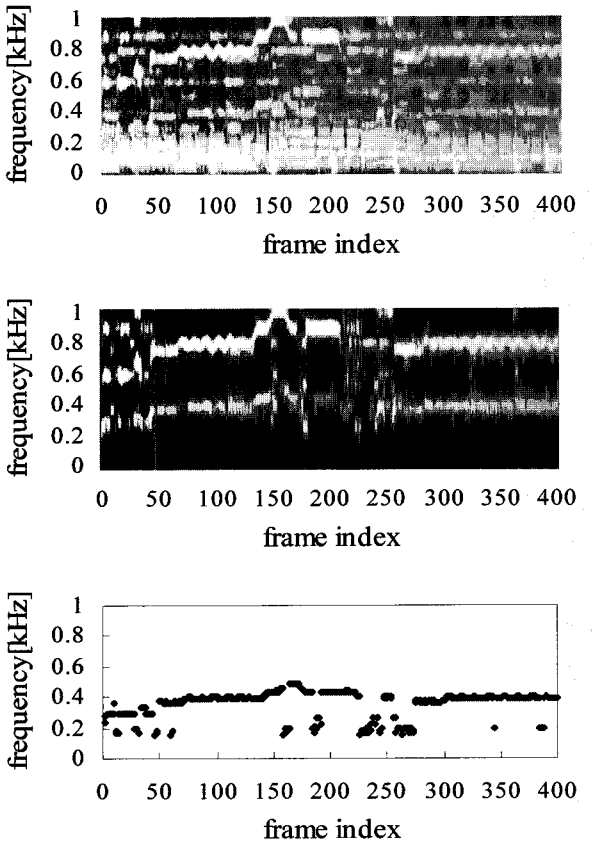

그림 6. 윤도현의 "잊을게"에 대한 멜로디 추출 결과(상단: 원 신호의 스펙트로그램, 중단: 검출된 멜로디의 스펙트로그램, 하단: 검출된 멜로디 주파수)

Fig. 6. Melody extraction result of Korean pop-song "잊을게" by Yoon Dohyun. (upper: original spectrogram, middle: spectrogram of extracted melody, lower: frequency of extracted melody).

500Hz 대역은 베이스, 드럼 등의 영향으로 기본 주파수 추출에 어려움이 있지만 본 논문에서는 전 대역에 대한 하모닉 성분을 추출하므로 효과적으로 멜로디 주파수를 추출함을 알 수 있다.

한국 가요에 대하여 정답 멜로디 주파수 정보가 제공되지 않으므로 RPA 로 성능을 측정하지 못하였다. 대신, 추출된 멜로디로 합성한 신호를 청취하여 원곡의 멜로디와 비교하여 멜로디 추출 성능을 평가하였으며, DB의 100곡에 대하여 주요 멜로디가 끊어지지 않고 매우 정확하게 표현되는 것을 확인하였다.

그림 6은 한국 가요인 윤도현의 "잊을게" 중 가사 "널 생각 하네 바보처럼" 구간에서 추출한 결과이다. 보컬이 포함된 프레임에서 멜로디가 정상적으로 검출되는 것을 볼 수 있다. 반면, 보컬이 없는 반주 영역에서는 음의 발생이 끊어지고 드럼을 포함하여 여러 악기들이 연주되어 검출 멜로디의 연속성이 감소됨을 알 수 있다. 그러나 이 영역은 주요 멜로디가 없는 구간이므로 제안한 멜로디 검출 방법의 성능이 저하된 것을 의미하지는 않는다.

Ⅳ. 결 론

본 논문에서는 다성 음악의 주요 멜로디를 검출하는 방법을 제안하였다. 하모닉 구조의 정확도를 기반으로 다중 피치를 검출하고, 검출된 기본 주파수 후보의 하모닉 에너지를 기반으로 중요도 순위를 정하였다. 피치 트래킹 모듈에서 기본 주파수 순위와 피치 연속성을 고려하여 최종 주요 멜로디를 검출하였다. ADC 2004 DB를 이용한 성능 측정을 통하여 제안한 방법이 기존 기술에 비하여 우수한 성능을 가지는 것을 확인하였다. 또한, 국내 가요에 대하여 스펙트로그램과 추출된 멜로디 신호에 대한 청취 평가를 통하여 제안한 방법이 주요 멜로디를 정확하게 검출하는 것을 확인하였다.

제안한 방법은 비교적 단순한 신호에 대하여 검출 성능이 우수하고, 복잡도가 높은 음악 또는 인위적으로 변조된 멜로디의 경우에는 정확도가 상대적으로 낮다. 또한, 특정 신호에 대하여 검출 정확도가 급격히 저하되는 경우가 발생한다. 이와 같은 문제점을 해결하여 모든 종류의 신호에 대하여 우수한 멜로디 검출 성능을 가지는 방법을 연구 중이다.

References

[1] 김무영, 이석필, "MIREX 기술 동향," 전자공학회지, 제37권, 제1호, 88-102쪽, 2010년 1월

[2] 박호종, 윤제열, "오디오 신호의 다중 피치 검출 기술," 전자공학회지, 제37권, 제1호, 63-72쪽, 2010년 1월

[3] M. Goto, "A robust predominant-F0 estimation method for real-time detection of melody and bass lines in CD recordings", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Process.*, Vol.2 pp.757-760, Istanbul, Turkey, June 2000.

[4] M. Goto, "A predominant-F0 estimation method for real-world musical audio signals: MAP estimation for incorporating prior knowledge about F0s and tone models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Process.*, pp. 3365-3368, Aalborg, Denmark, June 2001.

[5] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech and Audio process.*, Vol.11, No.6, pp.804-815, 2003.

[6] A. P. Klapuri, "Multiple fundamental frequency estimation by summing harmonic amplitudes," in *Proc. 7th Int. Symposium on Music Information Retrieval*, pp.216-221, Victoria, Canada, Oct 2006.

[7] M. Lagrange, L. G. Martins and J. Murdoch, "Normalized cuts for predominant melodic source separation," *IEEE Trans. Audio, Speech, Language process.*, vol. 16, no. 2, Feb. 2008.

[8] Y.-G. Zhang and C.-S. Zhang, "Separation of music signals by harmonic structure modeling," *Neural Information Processing Systems*, pp. 184-191, 2005.

[9] J.-L. Durrieu, G. Richard, and B. David, "Singer melody extraction in polyphonic signals using source separation methods," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Process.*, pp.169-172, Las Vegas, U.S.A. April 2008.

[10] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and inharmonic non-negative matrix factorization for polyphonic pitch transcription,".in *Proc. IEEE International Conference on Acoustics, Speech and Signal Process.*, pp.109- 112, Las Vegas, U.S.A. April 2008.

[11] E. M. Voorhess, D. M. Tice, "The TREC-8 Question Answering Track Evaluation," in *Proc. 8th Text Retrieval Conference*, pp. 77-82, NIST,

Gaithersburg, MD, 1999.
[12] G. Poliner, D. P. Ellis, A. F. Ehmann, E. Gomez, S. Streich, B. Ong, "Melody Transcription from Music Audio: Approaches and Evaluation," *IEEE Trans. Audio, Speech and Language Process.*, Vol. 15, No.4, pp.1066-1074, May 2007.
[13] http://www.music-ir.org/mirex/2009/index.php/ Audio Melody Extraction Results.

저 자 소 개

윤 제 열(학생회원)
2007년 광운대학교 전자공학과 학사 졸업.
2007년~현재 광운대학교 전자공학과 박사과정.
<주관심분야 : 음성/오디오 신호처리, 멀티미디어 신호처리>

이 석 필(정회원)
1990년 연세대학교 전기공학과 학사 졸업.
1992년 연세대학교 전기공학과 석사 졸업.
1997년 연세대학교 전기전자 공학과 박사 졸업.
1997년~2002년 대우전자 영상 연구소 선임연구원.
2002년~현재 전자부품연구원 디지털미디어연구센터 센터장.
<주관심분야 : 디지털방송통신융합시스템, 멀티미디어 신호처리>

서 경 학(정회원)
1978년 연세대학교 전자공학과 학사 졸업.
1980년 한국과학기술원 전자공학과 석사 졸업.
1989년 Syracuse Univ. 전기 및 컴퓨터공학과 박사 졸업.
1980년~2001년 삼성전자 Personal Multimedia 사업팀장.
2001년~현재 전자부품연구원 정보통신미디어연구본부장.
<주관심분야 : 무선통신, WPAN>

박 호 종(정회원)-교신저자
1986년 서울대학교 전자공학과 학사 졸업.
1987년 Univ. of Wisconsin-Madison 전기 및 컴퓨터 공학과 석사 졸업.
1993년 Univ. of Wisconsin-Madison 전기 및 컴퓨터 공학과 박사 졸업.
1993년~1997년 삼성전자 선임연구원.
1997년~현재 광운대학교 교수.
<주관심분야 : 음성/오디오 신호처리, 멀티미디어 신호처리>