

오디오 신호의 다중 피치 검출 기술

박호종·윤제일 (광운대학교)

I. 서 론

오디오 신호를 분석하여 신호에 포함되어 있는 음악의 구성 성분을 추출하고, 음악 내용을 이해하고, 음악 내용을 효과적으로 표현하는 기술 연구가 활발히 진행되고 있다. 이와 같은 기술의 대표적인 응용분야의 예로, 음악 신호의 자동 악보 변환(music transcription), 악기 인식, 보컬(vocal) 멜로디 추출, 내용 기반 음악 검색, 음악 장르(genre) 분류, 음악 감성 인식, 음악 내용 표현 등이 있다. 이와 같은 기술 분야에서 오디오 신호에 포함되어 있는 오디오 특성(feature) 파라미터를 검출하는 것이 핵심이고, 대표적인 특성 파라미터로 피치(pitch), 리듬(rhythm), 템포(tempo), 시간(temporal) 강약, 음색(timbre) 등이 있다. 이와 같은 특성이 정확히 추출되면 이 특성들을 조합하고 모델링하여 음악의 내용을 이해하고 분석할 수 있게 된다.

본 논문에서는 가장 대표적인 오디오 특성 파라미터인 피치를 검출하는 기술을 소개 한다. 피치 검출을 통하여 음악의 핵심 정보인 멜로디의 음 높낮이 변화를 추출할 수 있으며, 리듬 및 템포 정보와 결합하면 완전한 멜로디 정보를 얻도

록 해준다.

일반적인 오디오 신호는 보컬을 포함하여 다양한 악기에 해당하는 음원(sound source)이 혼합된 다성 음악(polyphonic music) 신호이다. 즉, 서로 다른 멜로디를 표현하는 여러 신호가 혼합된 결과이며, 이는 하나의 음원에 의한 단음(single note) 멜로디만을 가지는 단성 음악(monophonic music) 신호와 구별된다. 다성 음악 신호에서는 각 시간 영역에 여러 개의 서로 다른 음이 동시에 나타나므로 각 음에 해당하는 피치를 모두 검색하는 것이 핵심이다. 이와 같이 다성 음악 신호에서 여러 개의 피치를 검출하는 기술을 다중 피치(multi-pitch) 검출 기술이라 한다.

기존에 단성 음악 신호의 피치를 검출하는 기술이 많이 연구 되었고, 주로 음성 신호의 피치 검색을 위한 기술이 개발 되었다^[1]. 대표적으로 시간 영역 자기상관관계 (autocorrelation), HPS (harmonic product spectrum), 모렐 기반 방법 등이 있다. 그러나 이 기술들을 다성 음악 신호에 그대로 적용하면 원하는 다중 피치를 검출 할 수 없다. 따라서 다성 음악의 고유 특성을 고려하여 여러 개의 피치를 동시에 검출하는 새로

운 다중 피치 검출 기술에 대한 연구가 필요하다.

본 논문에서는 다성 음악 신호의 다중 피치를 검출하는 기술들을 소개한다. 다양한 기술들을 접근 방식에 따라 분류하여 설명하고, 대표적인 기술의 다중 피치 검출 성능을 간단히 정리한다.

II. 다중 피치 검출 기술

오디오 신호의 다중 피치 검출 기술을 Auditory Scene Analysis 모델, Auditory 필터 모델, 신호 확률 모델, 그리고 그 외 방법 등 4가지 접근 방식으로 분류하여 설명한다.

1. Auditory Scene Analysis(ASA) 모델

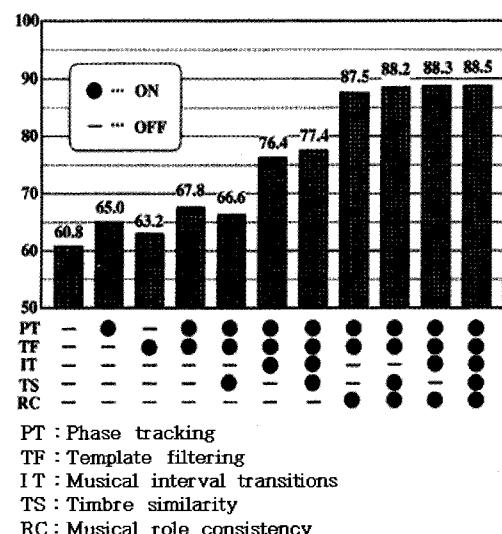
인간의 청각은 다성 음악을 청취할 때, 신호를 각 음원으로 분리하여 각 음원별로 인식하고 이해할 수 있는 능력을 가진다. 이와 같은 청각의 인지 능력을 ASA라 한다. 만일 다성 음악 신호를 수학적으로 분석하여 각 음원별로 분리시킬 수 있으면, 기존의 단독 피치 검출 기술을 각 음원 신호에 적용하여 각 음원에 해당하는 다중 피치를 검출할 수 있다. 이와 같은 배경에 따라 ASA 모델을 기반으로 다중 피치를 검출하는 기술이 연구되었다.

ASA의 구조 및 메커니즘은 Bregman^[2]에 의하여 제시되었고, 그 이후에 CASA (computational auditory scene analysis)로 발달하게 되었다. CASA는 일반적으로 2단계의 처리과정으로 구성된다. 첫 번째 단계는 음악 신호의 시간 및 주파수 성분들을 분리하는 과정이고, 두 번째 단계는 이렇게 분리된 성분들 중 동일한 음원으로 판단되는 성분들을 집단화시키는 과정이다.

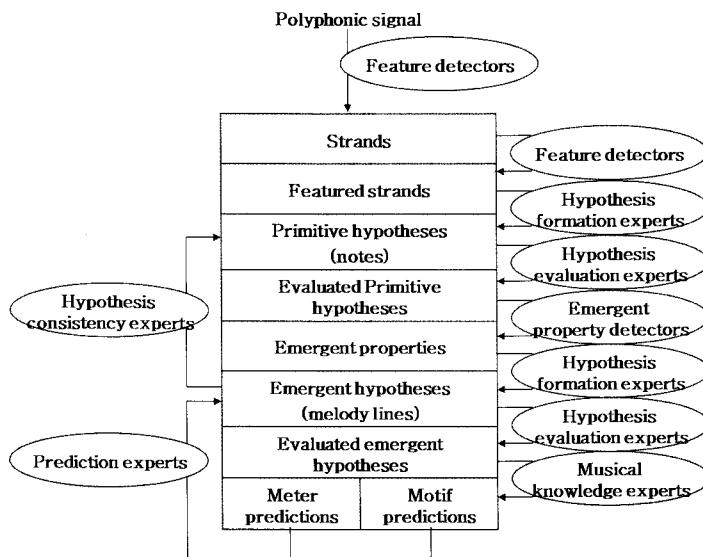
Kashino는 Bregman 이론에 따라 MSA (music scene analysis) 기술을 개발하였다^[3]. 음악 신호에서 시간에 따라 변하는 정현파 (sinusoidal) 성분을 검출하고, 새로운 음의 시작점과 하모닉(harmonic) 구조 등을 기반으로 정현파 트랙(track)을 집단화 하여 음표를 구한다. 또한, 음색(timbre)을 모델링하여 악기의 음색을 판별하려 하였고, 화음을 모델링하였다. 구체적으로, <그림 1>에서 보듯이 5가지 방법을 적용하였고, 3가지 악기를 합성한 신호를 입력 모델로 사용하였다. 5가지 방법을 모두 적용하였을 경우에 가장 좋은 성능을 가지는 것을 알 수 있다.

Sterian은 Kalman 필터를 사용하여 정현파 구성원을 추출하였고, 추출된 정현파 구성원에 대하여 Likelihood Function을 적용하여 집단화 하였다^[4].

Recognition rate [%]



<그림 1> Kashino가 제안한 방법의 성능
(from^[3], Fig.12)



〈그림 2〉 Godsmark과 Brown01 제안한 방법^[5]

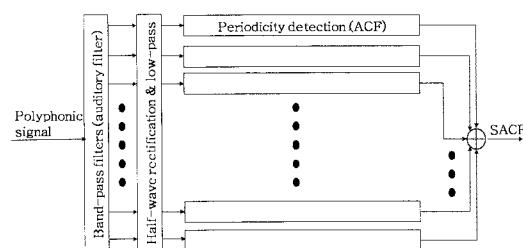
Godsmark과 Brown은 CASA 모델링과 음악적 규칙을 결합하여 음악적 특성에 따른 집단화 경계값을 적용하였다^[5]. <그림 2>에서 보듯이 여러 단계에 걸쳐 음악 특징을 추출하고 그와 관련된 음악적 규칙을 검증하여 집단화 경계값을 구한다. 또한 입력 신호의 피치와 음색 특징을 기반으로 신호를 분류하고 있는 것을 알 수 있다.

2. Auditory 필터 모델

인간의 청각 필터를 모델링하여 다중 피치를 검출하는 기술이 많이 개발 되었다. Meddis와 O'Mard는 청각 모델을 보다 적극적으로 활용하고 수학적으로 적용하여 “Unitary” 피치 모델을 제안하였다^[6]. 이들이 제안한 방법은 <그림 3>과 같이 구성된다. 인간의 청각 필터를 모델링하는 대역통과 필터뱅크(filterbank)를 이용하여 입력 신호를 대역으로 분해하고, 각 대역 신호의 포락선(envelope)의 주기성을 분석 하는 과정

으로 구성된다. 구체적인 과정은 다음과 같이 정리된다.

- 40~120개의 대역 통과 필터를 이용하여 입력 신호를 대역으로 분해
- 각 대역 신호를 반파 정류(half-wave rectification)와 저역 통과 필터링하여 대역별 포락선 구함
- ACF(autocorrelation function)을 적용하여 각 대역의 주기성 추정
- 각 대역의 ACF 값을 합하여 SACF (summary

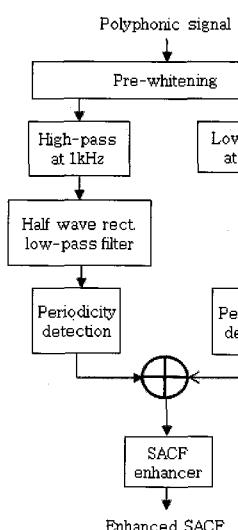


〈그림 3〉 Meddis와 O'Mard가 제안한 방법^[6]

autocorrelation function) 값을 정의하고, 이 중에서 가장 큰 값을 피치 값으로 정함

위에서 보는 바와 같이 Unitary 피치 모델 방식은 간단하면서도 우수한 성능을 제공하여 이후 많은 연구자들이 다중 피치 검출에 적용하였다. Cheveigné와 Kawahara는 Unitary 피치 모델 방식을 다중 피치 검출 영역으로 확대 적용하였고, 추출된 피치에 해당하는 성분을 신호에서 반복적으로 제거하여 다중 피치를 순차적으로 검출하는 방법을 제안하였다^[7]. 이와 같은 피치 검출 및 제거 방법은 그 이후에도 많은 연구자들이 사용하고 있다.

Tolonen과 Karjalainen은 Unitary 피치 모델 방식의 효과적인 계산 방법을 제안하고, 기존의 40~120개 대역 대신에 <그림 4>와 같이 2개 대역만을 사용하였다^[8]. 또한 신호에 포함되어 있는 잡음(noise)에 강인하게 동작하기 위한 전 처리 과정으로서 입력 신호를 역 LPC (linear predictive coding) 필터링 하여 주파수 특성을



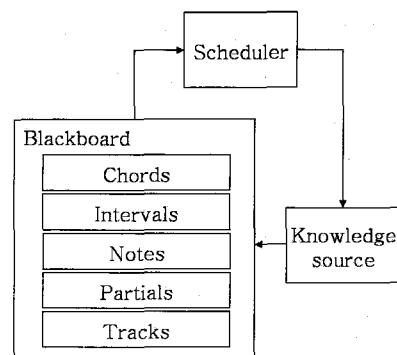
<그림 4> Tolonen과 Karjalainen이 제안한 방법^[8]

평坦하게 만들었다. 각 대역의 주기성 분석은 일반적인 ACF를 사용하였지만, 반복적인 피치 성분의 제거 없이 다중 피치 정보를 추출할 수가 있다.

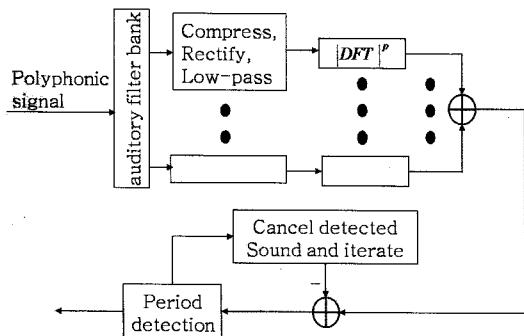
Martin은 <그림 5>와 같이 청각 모델 방식과 음악적 규칙을 통합적으로 사용하는 방식을 제안하였다^[9]. 청각 모델을 전처리기로 사용하고, 스케줄러에 의해서 신호의 음악 정보가 쌓이게 된다. 프레임 사이에 새로운 기본 주파수가 발생할 경우에 비교적 긴 시간의 포락선 정보를 사용하여 이전 기본 주파수의 정보에 따라 악보를 작성하게 된다. 이때 음악적 규칙을 적용하여 기본 주파수를 선별한다.

위에서 언급한 방법들에 적용된 청각 필터 모델은 일반적으로 사용되는 심리음향 데이터를 기반으로 설계되었다. 그러나 이러한 방법을 다성 음악 신호에 적용하는데 문제점이 있다는 것을 인식한 Klapuri는 <그림 6>과 같이 시간 영역에서의 ACF를 사용하지 않고 주파수 영역에서의 주기성, 즉 기본 주파수를 계산하는 새로운 방법의 Auditory 모델 방식을 제안하였다^[10].

[10]에서 제안한 방법은 Auditory 필터 모델링과 반복적인 기본 주파수 제거 방식을 결합한 구조이다. Auditory 필터는 일반적인 Gammatone



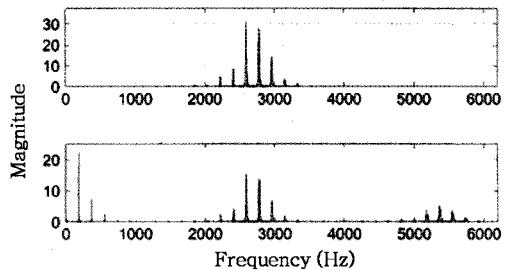
<그림 5> Martin이 제안한 방법^[9]

〈그림 6〉 Klapuri가 제안한 방법^[10]

필터를 사용하였고, ERB (equivalent rectangular bandwidth)에 따라 대역을 나눈다. 이 방법을 [6]에서 제안한 Unitary 피치 모델과 비교하면 대역 통과 필터링, 반파 정류, 그리고 저역 통과 필터링 과정까지는 동일하나, 주기성을 구하기 위하여 ACF을 사용하지 않고 DFT를 통하여 주파수 영역에서 주기성을 판별하는 큰 차이점을 가진다. 이는 기존의 ACF 방법이 잡음에 약하고 다성 음악에 적합하지 않기 때문이다. 즉, ACF를 적용한 기본 주파수 판별은 600Hz 이상의 기본 주파수 검출이 어렵지만, 주파수 영역에서는 40Hz ~ 2.1kHz의 기본 주파수를 판별할 수 있게 되고, 이에 따라 ACF를 이용한 방법에 비하여 고주파의 음원도 찾을 수 있다.

[10]의 방법은 각 대역별 스펙트럼을 이용하고 [14]와 동일한 방법을 적용하여 기본 주파수로 추출한다. 이 때, Inharmonic인 경우를 고려하여 대역별 최대값을 구하고 그에 따라 가중치 값을 결정한다. 이와 같은 방법으로 기본 주파수를 구한 후, 반복적으로 원음에서 기본 주파수를 제거하는 방식으로 다중 기본 주파수를 순차적으로 구한다.

〈그림 7〉은 HWR(half-wave rectification)에 의한 스펙트럼 변화의 예를 보여준다. 기본 주

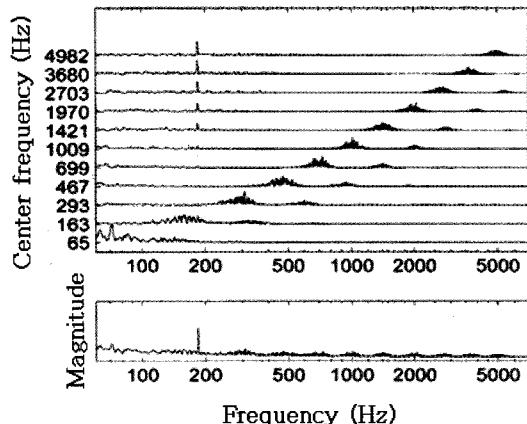


〈그림 7〉 HWR 적용에 의한 스펙트럼 변화

(upper: 원 스펙트럼, lower: HWR 적용 후 스펙트럼) (from^[10], Fig 4)

파수가 185Hz인 신호를 중심 주파수 2.7kHz인 대역통과 필터를 통과시킨 경우를 보여주며, 이를 통하여 HWR를 적용하여 하모닉의 특성이 뚜렷하게 나타남을 확인 할 수 있다.

〈그림 8〉은 Inharmonic 상황에서 Auditory 모델의 역할과 주파수 영역에서의 피치 검출의 장점을 구체적으로 보여준다. 백색 잡음과 185Hz의 단일 신호를 합성한 후 Auditory 필터 모델을 적용한 신호의 대역별 스펙트럼을 보여준다. 저 대역에서는 잡음 성분이 강하지만 고대역에서는



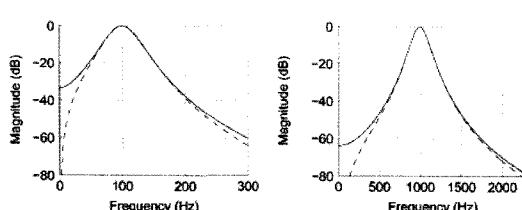
〈그림 8〉 백색 잡음과 단일 신호를 합성한 후 HWR을 적용(upper: 대역에 따른 스펙트럼, 각 대역을 합한 스펙트럼)

(from^[10], Fig 6)

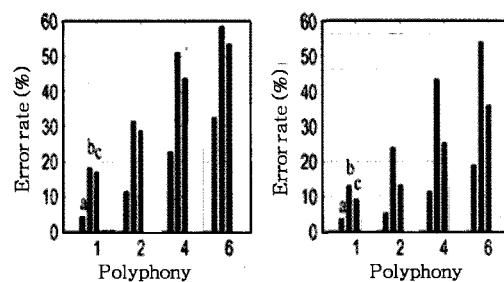
뚜렷한 185Hz의 단일 신호가 있음을 확인할 수 있다. 이를 통하여 Inharmonic인 상황에서도 기본 주파수를 강인하게 추출할 수 있으며, ACF 대신에 주파수 영역에서 주기성을 추출하는 것이 더 효과적임을 알 수 있다.

Auditory 필터에 주로 사용하는 Gammatone 필터의 부엽(side lobe) 특성이 다중 합성 음악 특성에 맞지 않다고 판단하여 [10]에서는 부엽이 '0'에 가까운 주파수 응답을 가지는 새로운 필터를 제안하였다. 즉, 음악 신호의 특성에 따라 중심 주파수에서 멀어질수록 감쇠가 강하게 일어나는 필터를 사용한다. <그림 9>는 [10]에서 사용하는 Gammatone 필터의 주파수 응답을 보여준다. 실선은 일반적인 Gammatone 필터이고 점선은 [10]에서 사용하는 새로운 필터의 모양이다.

<그림 10>은 [10]과 이전에 제안된 방법들의 다중 피치 추출 성능을 비교한 것이다. a는 [10]의 성능을 나타내고, b와 c는 각각 [8]과 [11]의 성능을 나타낸다. 왼쪽 그래프는 프레임 길이가 46ms, 오른쪽은 프레임 길이가 93ms인 경우를 나타내고, 두 경우에서 모두 [10]의 성능이 [8] 및 [11]에 비하여 우수한 것을 보여준다. 가로축의 Polyphony는 다성 음악에 포함되어 있는 음원의 수를 나타내며, 음원의 수가 증가할수록 성능이 저하되는 것을 볼 수가 있다.



<그림 9> Gammatone 필터 (100Hz, 1kHz)
(from^[10], Fig 3)



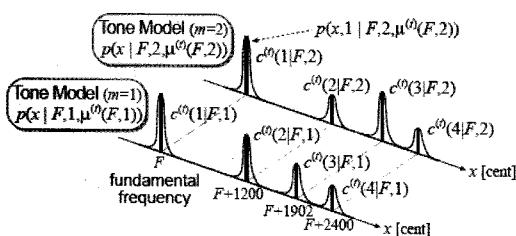
<그림 10> 다중 피치 검출의 성능

(left : 46ms frame, right : 93ms frame)
(from^[10], Fig 7)

3. 신호 확률 모델

음악 신호는 매우 구조적인 신호이므로 다중 피치 검출을 신호의 확률 모델을 기반으로 수행할 수 있다. [12]에서는 시변(time-varying) 진폭, 이상적이지 않은 하모닉 구조, 잡음 등을 적용하여 좀 더 정교한 신호 모델링을 제안하였다. 신호 모델을 파라미터화 하여 Likelihood Function에 적용하여 예측 값을 판별하였다. 이 때, 파라미터는 과거 값의 확률적 분포를 기반으로 시간 축에서 결정하고, 의미 있는 과거의 값을 추출하기 위하여 MCMC(Markov chain Monte Carlo)를 일정한 공간의 제한 없이 적용하였다. 그렇기 때문에 파라미터의 공간이 방대해지고, 과거의 확률 분포를 얻기 위하여 많은 연산량이 필요한 문제점이 있다.

Goto는 <그림 11>과 같이 톤(tone) 모델들을 만들어 서로 다른 가중치를 주고 합하여 음악 신호를 만들고 실험 하였다^[13]. 톤 모델은 우선 기본 주파수를 정수배 하여 하모닉 구조를 만든 후 Gaussian 분포를 적용하여 정의 하였다. 톤 모델을 반복적으로 업데이트하는 효율적인 방법을 제안하였으며, 일반적으로 예측에 사용되는



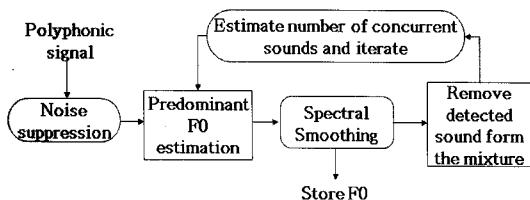
〈그림 11〉 Goto가 제안한 톤 모델들
(from^[13], Fig. 3)

MAP (maximum a posteriori) 방법에 EM (expectation maximization) 방법을 적용하여 편리한 연산 방법을 제안하였다. 또한 다중 연결 (multiple tracking)을 위하여 시간적 연속성을 고려하여 설계하였다.

4. 그 외의 방법

Klapuri는 주파수 영역에서 이상적인 하모닉 구조의 성질, 즉 기본 주파수의 정수배에 하모닉 피크가 나타나는 성질을 이용하고, 반복적으로 강한 피치를 구하고 제거하는 방법을 <그림 12>와 같이 제안하였다^[11]. 제안한 방법은 총 5단계로 이루어지고 모든 연산은 주파수 영역에서 이루어진다.

먼저, 전처리 과정으로 입력 신호의 잡음 성분을 제거하고, 주파수 영역에서 가장 강한 피치 값을 검출한다. 기존의 피치 추출은 시간 또는 주파수 영역에서 여러 개의 대역으로 나눈 후 각



〈그림 12〉 Klapuri가 제안한 방법^[11]

채널 별로 포락선의 주기 성질을 분석하여 피치를 찾는다. 그러나 기존 방법은 시간 영역에서는 피치 Halving, 주파수 영역에서는 피치 Doubling의 고전적인 문제점을 가지고 있으며, 이러한 문제점을 해결하기 위하여 [11]에서는 신호를 주파수 대역으로 나누어 각각의 독립적인 기본 주파수를 구한다. 50Hz ~ 6kHz를 로그 함수에 따라 18개의 대역으로 나누며, 각 대역은 2/3-octave를 포함하고 있다. 대역 필터는 삼각형 모양의 주파수 응답을 가지는 신호를 사용하였다.

다음, 에너지가 가장 큰 대역의 주파수 성분을 이용하여 가중치 값을 결정하며, 각 대역별로 주파수 피크 위치를 찾아 기본 주파수로 정한다. 이러한 방법으로 강한 피치 성분을 구한 후 Smoothing 과정을 거치게 된다. Smoothing 과정은 추출된 피치 성분을 강화해주는 동작으로, 일반적으로 강한 피치 성분을 추출한 후 바로 잔여 신호에서 빼주는 방법을 반복적으로 실행하는데, 기본 주파수의 가중치를 서로 다르게 적용하여 잔여신호에서 빼주게 된다. 마지막으로, 반복적으로 기본 주파수를 제거하는 과정을 종료시키는 조건이 필요하며, 다성 음악의 음원의 개수를 측정하여 측정된 횟수만큼만 반복하여 동작하게 된다.

[11]의 방법은 다성 음악의 특성상 하모닉 구조가 불안정 할 수 있다는 점을 고려하여 설계되었고, 그에 대하여 기존에 없던 Smoothing 과정을 추가하여 Inharmonic 문제점을 극복하려 하였다.

Klapuri는 [11]을 기반으로 또 다른 향상된 방법을 제안한다^[14]. 이 방법은 기존의 기술들에 비하여 간단하면서도 효율적이고, Summing Harmonic Amplitudes를 기반으로 하며, 매우 우수한 성능을 가지는 대표적인 다중 피치 검출



기술이다. 먼저, 음악 신호의 음색(timbre) 정보를 제거하기 위한 주파수 평탄화 과정을 적용한다. 다음, 주파수 변환하여 스펙트럼을 구하고, 식 (1)에 따라 Salience $s(\tau)$ 값을 구한다.

$$s(\tau) = \sum_{m=1}^M g(\tau, m) |Y(f_{\tau, m})| \quad (1)$$

여기서, $Y(\cdot)$ 은 평탄화된 신호의 스펙트럼, $f_{\tau, m}$ 은 기본 주파수 후보자(candidate)의 m 번째 하모닉 성분의 주파수 값을 나타내고, $g(\tau, m)$ 은 m 번째 하모닉 성분에 적용되는 가중치이다. 만일 $f_{\tau, m}$ 이 신호의 하모닉 성분에 해당하면 해당 위치에 스펙트럼 피크값을 가지고, 그에 따라 $s(\tau)$ 은 매우 커지게 된다. 따라서 최대 $s(\tau)$ 값을 가지는 주파수를 검출하여 기본 주파수를 구한다. 다음, 검출된 주파수 성분을 원 스펙트럼에서 제거하여 잔여 스펙트럼을 구하고, 위의 과정을 반복하여 순차적으로 기본 주파수를 검색한다.

[11]과 [14] 방법은 모두 주파수 평탄화 후에 특정 함수값이 가장 큰 값을 기본 주파수 후보자를 추출하고, 이 과정에서 적절한 가중치를 적용하여 잔여 신호에서 반복적으로 빼주어서 기본 주파수를 구하는 공통점을 가진다. 이 때, 가중치를 정하는 방법에 따라 성능에 많은 차이가 나타나고, 두 방법은 가중치를 얻는 파라미터가 다르고 가중치 최대값을 구하는 알고리즘에 차이를 가진다.

Vincent, Bertin와 Badeau는 NMF (non-negative matrix factorization) 방법을 적용하여 스펙트럼에서 NMF 모델을 사용하는 방법을

제안하였다^[15]. NMF는 입력 신호를 Tuning 모델에 따라 하모닉 집단과 Inharmonic 집단으로 모델링 한다. 3개의 Tuning 모델과 2개의 배음(overtone) 모델을 결합하여 총 6개의 NMF 모델을 만들게 된다. 입력된 신호는 ERB를 통과하고 NMF의 특성에 따라 하모닉과 Inharmonic 모델링 된 NMF의 모델 파라미터를 적응적으로 반응하게 된다.

그 외의 POPI(position pitch estimation), 하모닉 집단화, HMM(hidden Markov model) 등 다양한 방법들을 사용하여 다중 피치를 검출하는 기술이 개발되었다.

III. 결 론

본 논문에서는 여러 음원이 혼합된 다성 음악 신호의 특징을 분석하기 위한 핵심 기술로서, 다중 피치를 추출하는 주요 기술들을 소개하였다. ASA 모델, 청각 필터 모델, 신호의 확률 모델을 기반으로 하는 방법, 그리고 주파수 영역에서 하모닉 특성을 이용하는 방법 등으로 구분되며, 각 방법의 간단한 동작 구조를 설명하였다. 본 논문에서 소개한 방법들은 복잡도가 심하지 않은 음악 신호, 즉 많은 종류의 악기를 포함하지 않은 음악일 경우 우수한 성능을 보이지만, 음악의 복잡도가 증가할수록 성능이 저하 된다. 그러므로 더 많은 연구를 통하여 높은 수준의 음악적 모델링, 음악적 규칙 모델링, 음악적 특성을 고려한 변형된 주파수 변환, 청각 동작의 심층적 분석에 관련된 뇌에 관한 연구 등을 기반으로 새로운 다중 피치 검출 기술의 연구가 필요할 것이다.

참고문헌

- [1] W. Hess, *Pitch Determination of Speech Signals*, Springer-Verlag, 1983.
- [2] A. Bregman, *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.
- [3] K. Kashino and H. Murase, "A sound source identification system for ensemble music based on template adaptation and music stream extraction," *Speech Commun.*, 27, pp.337-349, 1999.
- [4] A. Sterian, "Model-based segmentation of time frequency images for musical transcription," Ph.D. dissertation, Univ. of Michigan, 1999.
- [5] D. Godsmark and G. J. Brown, "A blackboard architecture for computational auditory scene analysis," *Speech Commun.*, Vol.27, No.3, pp.351-366, 1999.
- [6] R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoust. Soc. Amer.*, Vol.102, pp.1811-1820, 1997.
- [7] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Amer.*, Vol.111, No.4, pp.1917-1930, 2002.
- [8] T. Tolonen and M. Karjalainen, "A computationally efficient multi pitch analysis model," *IEEE Trans. Speech Audio Process.*, Vol.8, No.6, pp.708-716, 2000.
- [9] K. D. Martin, "Automatic transcription of simple polyphonic music: robust front end processing," MIT Media Lab. Percept. Comput. Sect., Tech. Rep. No.399, 1996.
- [10] A. P. Klapuri, "Multipitch analysis of polyphonic music and speech signals using an auditory model," *IEEE Trans. Audio, Speech and Language Process.*, Vol.16, pp.255-266, 2008.
- [11] 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.
- [12] M. Davy, S. Godsill and J. Idier, "Bayesian harmonic models for musical signal analysis," In J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.), *Bayesian Statistics VII*, Oxford: Oxford University Press, 2003.
- [13] 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. Workshop on Consistent and reliable acoustic cues for sound analysis*, Aalborg, Denmark, 2001.
- [14] A. P. Klapuri, "Multiple fundamental frequency estimation by summing harmonic amplitudes," in *Proc. 7th Int. Conf. Music Information Retrieval*, 2006.
- [15] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and nonharmonic nonnegative

matrix factorization for polyphonic pitch transcription,”. *Int. Conf. Acoust. Speech Signal processing (ICASSP)*, pp.109-112, 2008.

저자소개



박호종

1986년 2월 서울대학교 전자공학과, 공학사
1987년 12월 Univ. of Wisconsin-Madison, 전기컴퓨터
공학과, M.S.
1993년 5월 Univ. of Wisconsin-Madison, 전기컴퓨터
공학과, Ph.D.
1993년 9월~1997년 8월 삼성전자 선임연구원
1997년 9월~현재 광운대학교 교수
주관심 분야 : 음성/오디오 신호처리, 멀티미디어 신호처리



윤제열

2007년 2월: 광운대학교 전자공학과, 공학사
2007년 2월~현재 광운대학교 전자공학과 박사과정
주관심 분야 : 음성/오디오 신호처리, 멀티미디어 신호처리