

UCC 음원분류를 위한 연주악기 분류에 대한 연구

권 순 일[†] · 박 완 주^{**}

요 약

사용자가 직접 연주하여 제작한 콘텐츠에서 많이 사용되는 악기는 기타, 피아노, 그리고 바이올린 이다. 이중 기타와 피아노가 만들어 내는 오디오 신호의 특성이 비슷하여 구분하기가 어렵다. 하지만 시간에 따른 신호의 에너지 변화가 피크(Peak)들을 중심으로 서로 다른 패턴을 보이는 것으로 분석되었다. 누적 히스토그램을 이용하여 피크 존재 가능성의 확률적 분포를 구한 후, 피크를 중심으로 그 주변의 주파수 대역 별 에너지 변화 패턴을 통계적 방법으로 모델링하여 실험한 결과 피아노와 기타의 구분 성공률이 최고 14% 정도의 향상을 보였다.

키워드 : 손수제작물(UCC), 음원분류, 악기인식, 피크탐색

Musical Instrument Recognition for the Categorization of UCC Music Source

Kwon, Soonil[†] · Park, Wanjo^{**}

ABSTRACT

A guitar, a piano, and a violin are popular musical instruments for User Created Contents(UCC). However the patterns of audio signal generated by a guitar and a piano are too similar to differentiate. The difference between two musical instruments can be found by analyzing the frequency variation per each band near signal peaks. The distribution of probability on the existence of signal peaks based on Cumulative Histogram were applied to musical instrument recognition. Experiments with statistical models of the frequency variation per each band near signal peaks showed the 14% improvement of musical instrument recognition.

Keywords : User Created Contents(UCC), Musical Instrument Recognition, Peak Detection

1. 서 론

최근 손수제작물 (User Created Contents)에 대한 관심은 다양한 내용과 수준의 콘텐츠를 양산하는 촉매제 역할을 하고 있으며, 또한 일반인들이 각자의 취향이나 의도에 맞는 콘텐츠를 직접 제작할 수 있는 기회를 마련해 주고 있다. 이러한 사용자가 중심이 되어 제작되고 공유되는 UCC의 양이 방대해 지면서, 이를 효과적으로 관리하고 고속으로 정확하게 검색할 수 있는 기술이 요구되고 있다. 구체적으로는 콘텐츠, 식별정보, 소셜정보를 통합한 새로운 UCC 메타데이터 정의와 사용자의 관심, 취향 등을 분석하여 검색어 관련 정보 실시간 추천 등이 있겠다.

컴퓨터 네트워크를 수단으로 한 디지털콘텐츠의 대중화 속에 가장 눈에 띄는 카테고리 중 하나는 음악연주와 관련

된 음원 콘텐츠일 것 이다. 특히 사용자들이 자작으로 연주한 음원들이 많이 생성되고 있고, UCC란 형태로 검색의 주요한 대상이 되고 있다. 본 논문은 사용자가 검색하는 UCC들의 메타정보를 자동으로 검증해 보는 연구와 관련된 것인데, 특히 메타정보에 있어서 음악 분류 내에서의 메타정보를 입력하지 않았거나 부족할 경우, 특정음악의 콘텐츠를 검색하기가 쉽지 않고 검색자의 관심이나 취향을 분석해 내기 어려운데, 이러한 문제점을 극복하려 한다. UCC 등록 시 주관적 분류 및 메타정보의 객관적 검증을 통하여 보완적 정보를 추가한다면 검색의 신뢰성과 정확성을 향상시키고 사용자의 취향 분석에 도움이 될 것 이다.

UCC 콘텐츠 중에 사용자가 직접 연주한 음원에 있어서 가장 빈번하게 사용되는 악기들은 기타, 피아노, 바이올린 등이 있다. 기타와 바이올린은 현악기, 피아노는 건반악기로 분류된다. 현악기는 공명판 위에 매어 놓은 현의 진동을 발음체로 하는 악기의 총칭인데, 건반악기 중 피아노는 현악기로 분류되기도 한다. 현의 진동을 발생시키는 수단으로 손이나 도구를 사용하게 되는데, 기타는 손가락이나 피크로

[†] 정 회 원 : 세종대학교 디지털콘텐츠학과 교수
^{**} 정 회 원 : 한국과학기술연구원(KIST) 연구원
논문접수: 2009년 9월 1일
수 정 일 : 1차 2009년 11월 16일
심사완료: 2009년 12월 7일

현을 튕기는 방법으로 소리를 만들어내고, 바이올린은 활을 현에 문질러 소리를 내고, 피아노는 건반을 누르면 이와 연결되어 있는 나무망치가 현을 두드리는 방법으로 소리를 만들어 낸다. 넓게 보면 모두 현을 이용하여 소리를 만들어낸다는 공통점이 있다. 그런데 하나의 음표를 소리로 만들어 낼 때 바이올린의 경우 활과 현의 마찰이 연속적으로 일어나 그 소리 강도가 대체로 일정하다. 이에 비해 기타는 일반적으로 현을 튕기는 순간 소리가 강하고 그 이후 시간이 경과하면서 소리의 감쇄가 진행된다. 피아노의 경우도 현이 두들겨지는 순간 소리가 가장 강하고, 이후 점차 소리는 감쇄한다. 이러한 이유로 바이올린은 기타나 피아노와 다른 소리의 패턴을 보이지만, 기타와 피아노는 서로 흡사한 패턴의 오디오 신호를 가지고 있다.

UCC에 있어서 기타와 피아노는 가장 많이 사용되는 음악 연주 콘텐츠의 도구이지만, 실제로 오디오 신호만을 가지고 두 악기가 만들어낸 소리를 구분해 내기가 가장 힘든 조합 중에 하나일 것이다. 일반적으로 악기인식(Musical Instrument Recognition)에 사용되는 특징(Feature)들을 천편일률적으로 적용해서는 기타와 피아노를 구분하는데 한계가 있다. 이번 연구의 사전조사에서 기타와 피아노의 소리는 음표마다 소리가 발생하는 순간, 즉 현을 튕기거나 두드려지는 순간(피크, peak)을 중심으로 그 주변에서 서로 다른 특성을 갖고 있는 것으로 분석되었다. 그래서 피크탐색을 통하여 피크 주변의 오디오 신호만을 이용하여 기타와 피아노를 구분해 본 결과, 오디오 신호 전체에 일률적으로 적용했을 때보다 최고 14%의 악기인식률 향상을 이루어 냈다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 음악 연주악기의 인식과 관련된 주요연구들에 대해 소개하고 이번 논문의 연구와 비교해 본다. 제 3장은 이번 연구의 대상인 기타, 피아노, 그리고 바이올린의 오디오 신호적 특성과 이들을 구분하기 위해 사용된 방법에 대해 구체적으로 설명한다. 제 4장은 기타, 피아노, 바이올린을 구분하는 실험의 구성에 관하여 기술한 후, 제 5장에서 실험 결과에 대해 자세히 분석하고 유용성을 제시한다. 제 6장은 본 논문의 결론을 내린다.

2. 관련 연구

최근 10년간 연주악기에 대한 인식을 위한 연구가 이어져 오고 있다. K. Martin 와 Y. Kim은 14개의 오케스트라를 구성하는 악기들에 대한 인식을 위해 한 가지 톤으로 연주된 데이터를 가지고 MAP (Maximum a posteriori) 등의 패턴인식 방법을 적용해 보았고, 약 60에서 70%의 인식률을 보였다[1]. A. Eronen 와 A. Klapuri는 30개의 오케스트라를 구성하는 악기들에 있어서 각각 한 가지 악기로 연주된 데이터를 가지고 Short-Time RMS Energy Envelope, 즉 Rise Time, Decay Time, Strength of Amplitude Modulation, Crest Factor, Exponential Decay 등 43가지의 특징(Feature)들을 이용하여 악기 인식을 시도해 보았고, 개개의 악기인

식에 있어서 약 80%의 인식률을 보였다[2]. H. Hacihabiboglu 와 N. Canagarajah는 웨이블릿(Wavelet)을 이용하여 플루트(Flute), 트럼펫(Trumpet), 그리고 클라리넷(Clarinet)을 구분하는 실험을 하였으며, 약 78%의 인식률을 나타냈다[3]. S. Essid 등 3명은 Mel-scale Frequency Cepstral Coefficient(MFCC) 등의 특징을 가지고 Support Vector Machine(SVM)의 방법을 이용하여 금관악기, 목관악기 등 10개의 악기인식을 시도했고, 그 결과 평균적으로 87%의 인식률을 보였다[4, 6]. A. Livshin 와 X. Rodet은 7가지 악기(Bassoon, Clarinet, Flute, Guitar, Piano, Cello, Violin)를 20가지 특징(Spectral Slope, Spread, Loudness, MFCC, Sharpness 등)들로 85%의 인식률을 기록했다[5]. T. Kitahara 등 5명은 피아노, 바이올린, 클라리넷, 플루트 등 4개의 악기를 Spectral and Temporal Feature로 이루어진 28개의 특징들과 Hidden Markov Model을 이용한 통계적인 모델링을 이용하여 인식을 실험결과 약 85%의 정확도를 보였다[7]. A. Cont 와 S. Dubnov는 피아노와 바이올린으로 연주된 곡들 섞어서 동시에 두 개의 악기로 연주된 데이터로부터 피치정보를 가지고 두 악기를 구분해 내는 실험을 하였으며, 그 결과 피아노는 53%, 바이올린은 89%의 인식률을 이루어 냈다[8]. E. Hassan은 신경망모형(Neural Networks)을 이용하여 19개의 오케스트라를 구성하는 금관, 목관, 현악기들을 구분해 보려는 시도를 했다[9].

위에서 본 것과 같이 지금까지의 연구는 주로 다수의 악기를 인식하기 위해 많은 가능성이 있는 특징들을 차별화된 선택적 적용 없이 사용하였고, 평균적으로 80% 대의 인식률을 보여 왔다. 하지만 이 논문에서는 UCC와 관련하여 음원 분류에 필요하면서도 지금까지 시도되지 않았던 기타와 피아노를 구분해내는 방법을 시도하였고, 특히 두 가지 악기의 특징을 구별할 수 있는 최선의 방법을 찾아내는데 있어서 피크탐색을 통해 선택적으로 오디오 신호 데이터를 활용하였다는 데에 차별성이 있다.

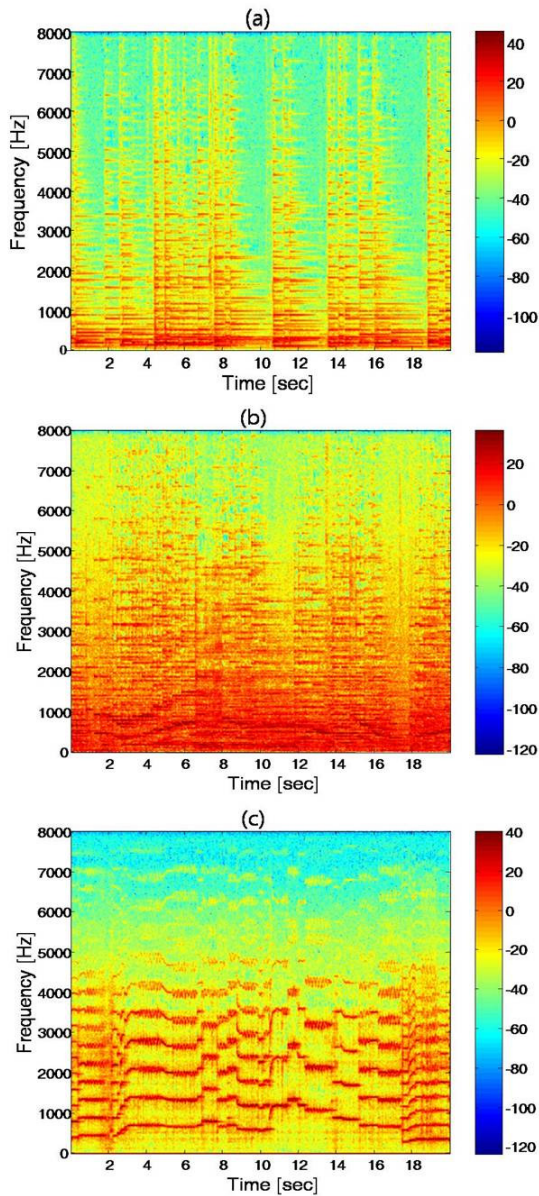
3. 기타, 피아노, 바이올린의 오디오신호 분류

3.1 오디오신호 패턴들

대체로 UCC로 만들어진 일반인들의 솔로 연주곡에서 가장 많이 사용되고 있는 악기는 기타, 피아노, 그리고 바이올린이다. 이들 악기들 중 피아노를 제외하면 나머지 두 악기는 현악기에 속한다. 현(줄)을 진동시켜 소리를 내는 악기를 통틀어 현악기라고 한다. 현악기의 음색은 최고 음역에서 최저 음넓이에 이르기까지 같은 음질이며, 여러 가지 다양한 음향을 낼 수도 있다. 현악기는 대개 연주방법에 따라 두 가지 종류로 나눌 수 있다. 활로 현을 문질러서 소리를 내는 악기와 손가락으로 튕겨서 내는 악기가 있다. 활로 현을 문질러 소리를 내는 악기에는 바이올린, 비올라, 첼로, 더블베이스가 있고, 손가락으로 튕겨서 내는 악기에는 하프, 기타 등이 있다. 이중 바이올린은 현악기 중에서 가장 높은 음을 내는 악기로 독주하기에 가장 알맞다. 기타는 6개의

줄을 가진 악기로 8자 모양의 공명 상자와 자루로 이루어져 있다. 왼손으로 현을 누르고 오른손으로 손가락이나 피크로 줄을 튕겨서 연주한다. 반면 피아노는 피아노포르테(Pianoforte)의 약칭이며 화성악기·선율악기의 두 요소를 갖추고 있는 건반악기로서 건반은 88건반이 표준이다. 음량은 풍부하고 남아있는 소리가 길며 센 음과 약한 음을 마음대로 낼 수 있다. 피아노는 건반과 연결되어 있는 나무망치가 현을 두드려 진동을 만들어 소리를 낸다. 이렇듯 바이올린, 기타, 피아노는 현악기이거나 건반악기로서 활을 사용하거나 손가락, 또는 나무망치 등 서로 다른 도구를 사용하지만, 현을 진동시켜서 소리를 만들어 낸다는 데에는 공통점을 가지고 있다.

(그림 1)은 기타(a), 피아노(b), 그리고 바이올린(c)으로



(그림 1) 각 악기별 독주 연주곡 음원 일부의 스펙트로그램: (a)기타, (b)피아노, (c)바이올린.

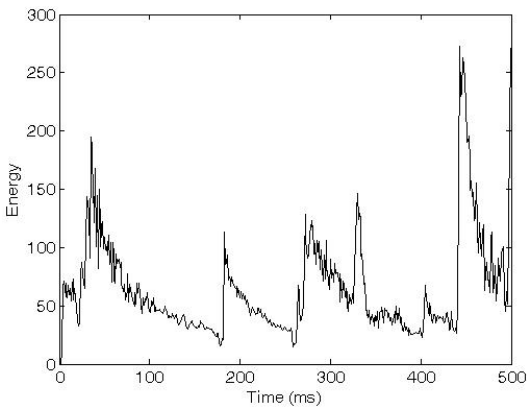
각각 연주된 음원의 스펙트로그램(Spectrogram)을 보여주고 있다. 각각 서로 다른 곡을 연주한 것이지만, 이 그림으로부터 각 악기가 주파수 대역에 있어서 어떠한 특징을 보여주는지 어느 정도 알 수 있다. 서로 다른 도구로 현을 진동시키고, 현의 진동에 의해 공기의 진동이 발생하고 이것이 서로 다른 모양과 크기의 울림공간을 통과하면서 다른 형태의 오디오 신호를 발생시킨다. 스펙트로그램 이미지로 볼 때 세 악기 중에 바이올린은 확연히 달라 보인다. 주파수 대역별로 오칠 모양으로 구분지어지고, 시간의 변화에 따른 음의 신호도 각 음표마다 강도 또는 음의 세기가 거의 일정하게 유지되는 것을 볼 수 있다. 이는 어떤 각 음표를 연주할 때 시작부터 끝까지 바이올린의 현이 활을 이용한 지속적 마찰을 통해 동일한 강도로 진동하기 때문이다. 반면 기타나 피아노는 손가락(또는 피크)으로 줄을 튕겨거나 나무망치로 현을 두드리는 방식이라 음표를 연주하는 시작시점에 소리가 최고로 크고 그 이후 점차 시간이 흐름에 따라 소리의 크기가 감쇄되는 모습을 보인다. 그래서 바이올린은 기타나 피아노와 구분이 수월한 편이지만 기타와 피아노는 구분이 쉽지 않다.

3.2 제안된 패턴분류 방법

이 논문에서는 UCC에서 가장 많이 사용되면서도 자동적인 방법으로 구분이 어려운 기타와 피아노 독주 연주곡 음원에 대한 분류가 가능하도록 하기 위해서 기타와 피아노로 만들어지는 오디오 신호의 패턴을 분석하려고 한다. 그림 1에서 볼 수 있듯이 기타와 피아노도 눈으로 볼 때는 어느 정도 구분이 되어 보이지만, 주파수 영역에서 평균적인 특징(Feature)을 대역별로 계산해서 관찰해 보면 에너지의 분포나 시간대비 에너지의 상승과 감쇄곡선의 패턴이 비슷한 양상을 보인다. 그래서 기존에 오디오 신호의 패턴, 특히 연주악기를 분류하기 위한 특징(Feature)으로 사용되어온, Rise Time (Duration of Attack), Slope of Energy Curve, Decay Time, Spectral Centroid, Fundamental Frequency 등을 일률적으로 적용해서는 기타와 피아노를 구분하기가 힘들다. 그런데 기타와 피아노 오디오 신호를 조금 더 자세히 분석해 보면, 오디오 신호 에너지의 시간 대비 피크에서와 그 주변에서 주파수대역 별 에너지 분포에서 약간의 차이가 있다. 피크의 순간에서 기타는 저주파영역에서 고주파영역으로 갈수록 에너지양이 고르게 감소하는 분포를 보이지만, 피아노는 대체로 고주파영역으로 가면서 감소하는 에너지의 분포가 불규칙적인 편이다. 또한 기타는 비교적 3-8kHz의 영역에서 속도만 다를 뿐 에너지의 시간대비 차연적 감쇄가 고르게 나타나는 반면, 피아노는 고르지 못한 편이다. 이러한 미세한 차이가 감지되지만 이러한 것들은 임의의 곡을 연주할 경우 연주곡에 따른 편차를 감안해 본다면 사실상 구분의 잣대로 사용하기가 용의하지 않다. 왜냐하면 기타의 현이 튕겨지는 시점, 그리고 피아노의 건반이 눌러지는 시점인 에너지의 시간적 피크를 정확하게 감지해 내고, 그 피크주변의 주파수대역 별 에너지의 변화를 관

찰해야 하는데, 피크를 감지하기가 쉽지 않다는 것이다. 이 논문에서는 가장 기본이 되는 오디오 신호 패턴분류 방법인 주파수 대역별 에너지의 변화를 비교 분석하는 방법을 기초로 하여 만일 피크를 감지하고 그 주변에서 발생하는 Rise Time, Decay Time과 관련된 변화의 패턴을 분석할 경우 얼마나 기타, 피아노, 그리고 바이올린의 연주를 구분하는데 도움을 주는지를 알아보려 한다.

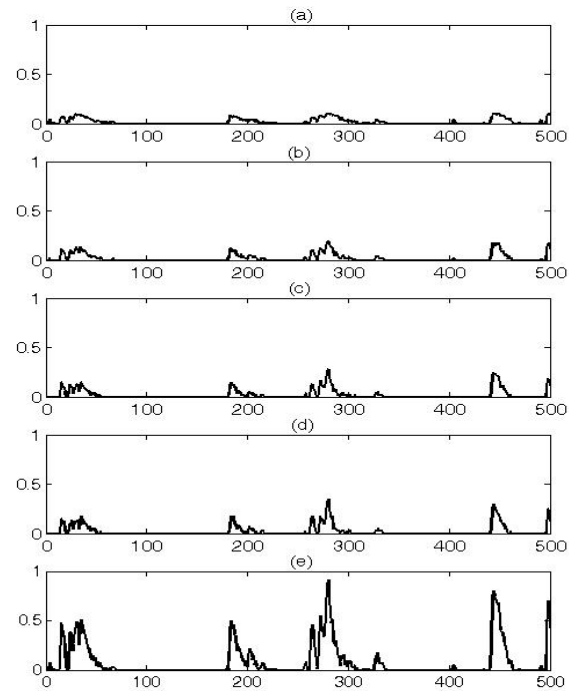
피크의 위치는 기본적으로 음악 악보의 각 음표 위치와 대응되지만, 임의의 연주곡이라는 것을 감안하자면, 피크들 사이의 간격이나 피크들의 크기, 즉 피크의 에너지 세기는 곡 전체를 연주하는 동안 연주기법과 곡의 특성에 의해 그 편차가 심할 수 있다. 그래서 단순한 한계값(Threshold)을 이용하는 것은 충분치 않다. 또한 전체주파수 대역을 여러 개의 밴드로 나누어 주파수 대역의 밴드 별 피크를 중심으로 그 주변의 변화를 관찰하는 것이 필요하다. 그 이유는 다음과 같다. 첫째, 에너지의 피크 이후 감쇄가 일어날 때, 저주파일수록 감쇄의 속도가 느리고 고주파일수록 감쇄의 속도가 빠르다. 둘째, 저주파의 감쇄속도가 느리기 때문에 피크와 피크사이의 간격이 좁을수록 피크를 찾기 힘들고 또한 감쇄의 패턴을 관찰하기 힘들다. 마지막으로 고주파 영역은 감쇄가 빨라 피크의 위치가 잘 나타나지만, 기본적인 에너지가 작아 항상 관찰이 용이하지는 못하다. 그래서 오디오 신호를 우선 밴드 별로 나누고, 각 밴드 별로 시간 대비 에너지 값을 누적 히스토그램(Cumulative Histogram)으로 계산해서 복수 개의 상위 특정 퍼센트 분위(Percentile)에 해당되는 값을 밴드 별 한계값으로 정하여 피크의 존재가능성을 확률적 분포로 만들어 보려 한다 [10, 11]. 예를 들어, (그림 2)의 기타로 연주한 음원의 오디오 신호에서 피크의 위치를 알기위해 우선 오디오 신호의 0Hz에서 8000Hz까지를 16개 밴드로 나눈 후 각 밴드 별 에너지 값으로부터 밴드 별 누적 히스토그램을 이용하여 상위 16% 값, 상위 8% 값, 상위 4% 값, 상위 2% 값을 구하여 각각의 밴드 별 한계값으로 설정한다. 이후 각 밴드 별로 한계값을 넘은 위치의 값은 원래 값에 가중치를 곱한다. 가중치는 위의 각 상위 분위 순서에 따라 '0.84', '0.92', '0.96', '0.98'과 비례하는 값으로 정한다. 한계값을 넘지 못한 위치의 값은 '0'을 곱한



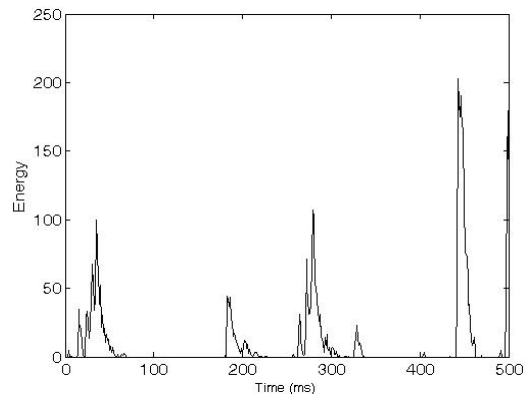
(그림 2) 기타연주 오디오 신호의 예

다. 마지막으로 모든 밴드의 결과를 합산한다. 그 결과로 (그림 3)을 얻을 수 있는데, 상위 16%(그림 3)(a)의 결과, 상위 8%(그림 3)(b)의 결과, 상위 4%(그림 3)(c)의 결과, 상위 2%(그림 3)(d)의 결과이고, (그림 3)(e)는 위의 4가지 경우를 모두 합하여 최종적으로 피크존재의 가능성을 확률로 표현한 결과가 된다. 이렇듯 한계값을 다양화 시키고 한계값과 비례하는 값을 가중치로 활용하며, 그 결과로 얻어진 확률 값을 원래의 샘플 값과 곱하여 피크의 주변 신호만을 남긴다. 또한 피크에 가까울수록 높은 가중치가 곱해지고, 피크에서 멀어질수록 낮은 가중치가 곱해지는 양상을 보이게 된다(그림 4).

지금까지 설명한 방법이 기존의 기초적인 밴드별 에너지



(그림 3) 누적 히스토그램 기반의 한계값을 이용한 피크의 존재 가능성 관련 그래프(단위: x축은 msec, y축은 확률값): (a)상위 16%, (b)상위 8%, (c)상위 4%, (d)상위 2%, (e)총합(a+b+c+d).



(그림 4) 기타연주 오디오 신호(그림 2)에 피크 존재 확률 값(그림 3)(e)을 곱한 결과

의 변화를 관찰하는 것과 다른 점은 피크를 중심으로 그 주변의 밴드 별 에너지 변화에 초점을 맞추는 선별적인 관찰을 한다는 것이다. 왜냐하면 기타와 피아노의 차이가 주로 피크를 중심으로 그 주변에서 나타나기 때문이다. 피크 주변의 신호 값만을 가지고 밴드별 에너지를 통계적으로 모델링하여 기타와 피아노를 구분하게 된다.

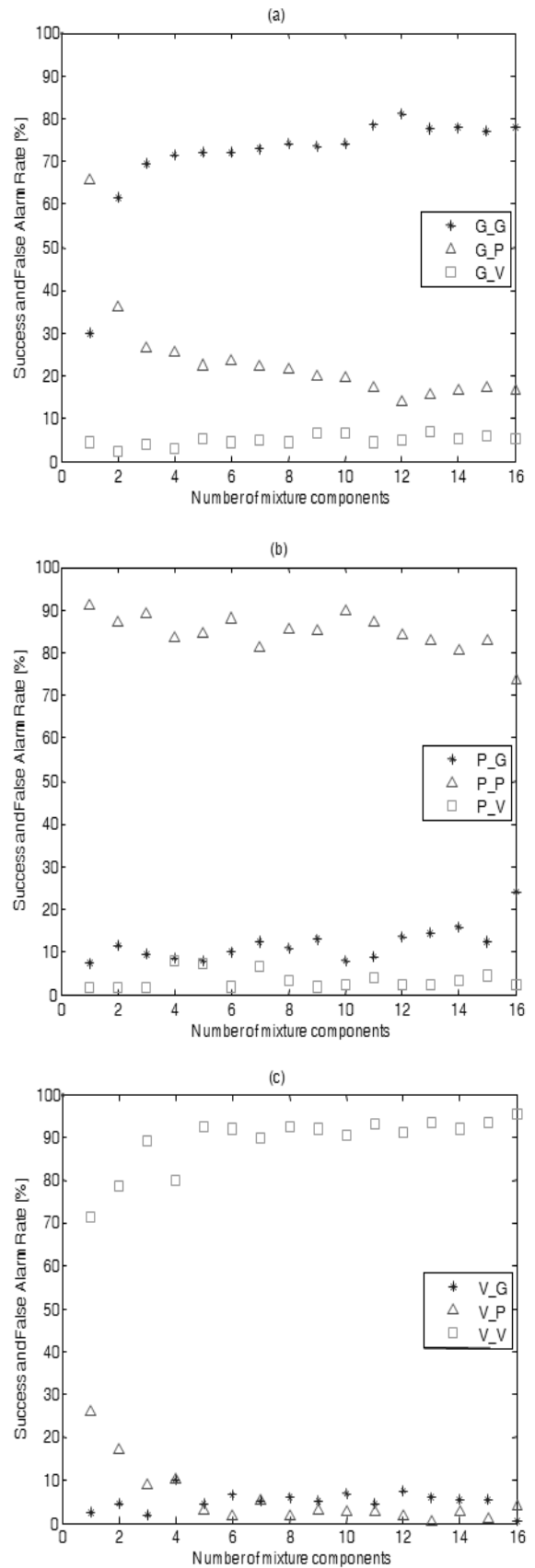
4. 실험

이 논문에서는 UCC에서 접할 수 있는 음악 연주곡들을 실험 데이터로 활용하였다. UCC의 음악연주곡을 찾기 위해 YouTube (www.youtube.com)에서 피아노, 기타, 그리고 바이올린 연주곡을 검색하였고, 그 중에서 각 악기 별로 단독 연주를 한 데이터 파일들을 선별하였다. 각 악기 별로 20곡을 선별하고 그 중에서 악기가 연주된 부분에 한하여 임의로 20초 길이의 데이터 파일을 만들었다. 즉 피아노 데이터 파일 20개, 기타 데이터 파일 20개, 바이올린 데이터 파일 20개를 실험에 이용하였다. 통계적 모델인 Gaussian Mixture Model을 이용하였는데, 훈련을 위한 데이터 파일과 테스트를 위한 데이터 파일을 구분하여 실험을 하였고, 훈련 역할과 테스트 역할을 서로 바꾸어 가며 반복해서 실험하였다. 각 연주곡의 곡명이나 연주 실력은 고려하지 않고 임의로 선택되었다. 그래서 피아노, 기타, 그리고 바이올린의 연주곡 중 동일한 곡을 연주한 곡은 없으며, 곡의 장르도 임의로 섞여 있다. 연주 데이터 파일은 형식은 웨이브 파일로 하였고, 샘플링 주파수는 16kHz로 변환하여 동일한 조건으로 만들었다. 대부분의 연주곡의 데이터 형식이 16bit Linear PCM 스테레오채널 형식이었지만, 실험 편의상 데이터 형식은 16bit Linear PCM 모노채널로 변환시켰다[12].

각 연주곡의 오디오 음원 데이터 분석을 위해 MATLAB이 사용되었다. 이를 이용하여 주파수 도메인에서의 스펙트럼 분석 및 특성 벡터 추출, 그리고 특성 벡터들을 이용한 모델링과 모델을 이용하여 패턴 분류를 통한 각 연주 악기 인식을 수행하였다. 주파수 도메인에서의 상세한 분석을 위해 0에서 8kHz까지를 500Hz 간격으로 16개 밴드로 구분하였다. 시간 도메인에서 윈도우 사이즈는 2초로 했고, 윈도우 이동은 2초씩 겹침 없이 했다. 윈도우 사이즈를 2초로 한 이유는 음악 연주 시에 보통 악보의 한 마디가 1.5초에서 2초 정도의 시간적 길이를 나타내기 때문이고, 이 논문에서는 한 마디를 패턴분석의 기본 단위로 채택했다. 모델링에 사용한 모델은 Gaussian Mixture Model로 Mixture의 수는 2개에서 16개 까지 다양하게 적용해 보았다.

5. 결과 분석

(그림 5)는 오디오 신호로부터 2초 길이의 윈도우를 통해 뽑아낸 데이터를 500Hz 간격으로 0-8000Hz를 16개 밴드로 나누고, 각 밴드 별 에너지를 구하여 이를 특징벡터

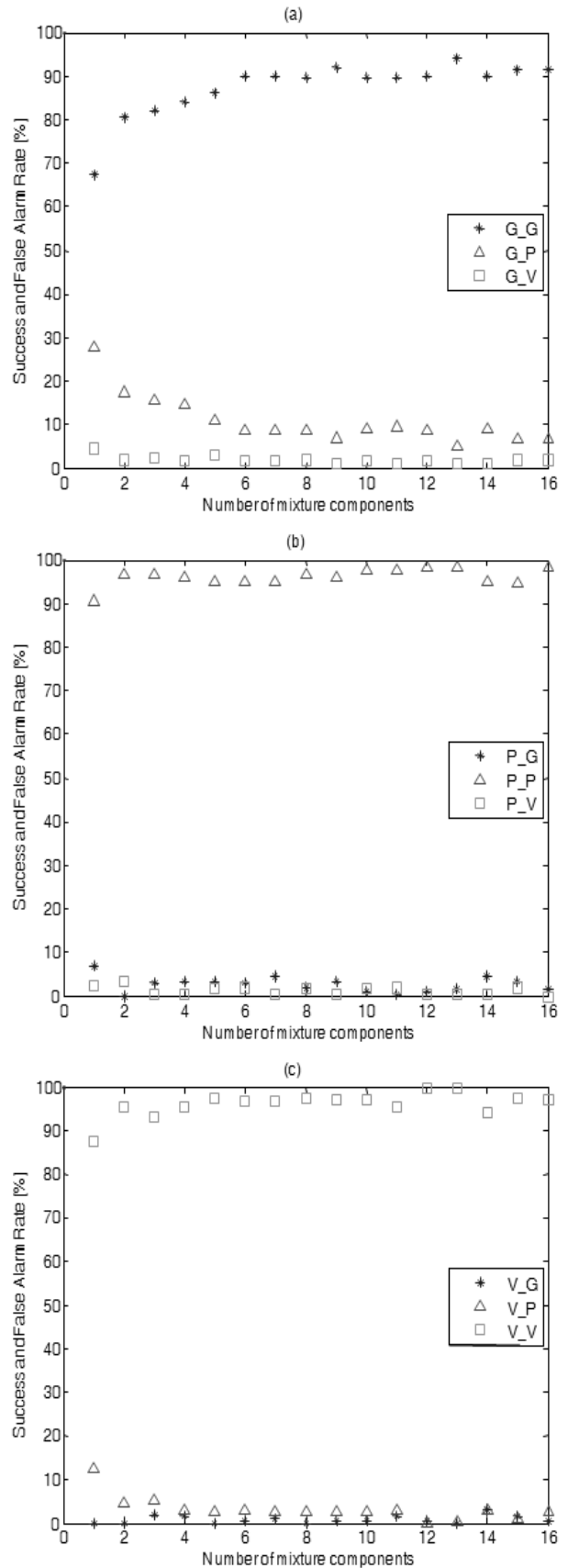


(그림 5) 주파수 대역 별 에너지 분포 기반 연주악기 인식을: 기타(G), 피아노(P), 바이올린(V)

(Feature Vector)로 하고, 특징벡터들로 훈련된 Gaussian Mixture Model을 이용하여 통계적으로 연주악기를 인식하는 방법을 실험한 결과이다. (그림 5)의 (a)는 기타 연주곡을 테스트 했을 때, 기타로 인식된 결과(G_G)와 기타가 아닌 피아노(G_P) 또는 바이올린(G_V)으로 인식된 결과를 보여준다. 기타의 Gaussian Mixture Model에서 Mixture 개수를 1부터 16까지 바꾸어가며 실험한 결과 최저인식률은 Mixture 수가 '1'일 때 30% 이었고, 최고인식률은 Mixture 수가 '12'일 때 82.5% 이었다. 대부분 바이올린보다는 피아노로 인식되는 오류의 빈도수가 많았다. (그림 5)의 (b)는 피아노 연주곡을 테스트 했을 때, 피아노로 인식된 결과(P_P)와 기타(P_G) 또는 바이올린(P_V)으로 인식된 결과를 보여준다. 최저인식률은 Mixture 수가 '7'일 때 80.5% 이었고, 최고인식률은 Mixture 수가 '9'일 때 90.5% 이었다. 대부분 바이올린보다는 기타로 인식되는 오류의 빈도수가 많았다. (그림 5)의 (c)는 바이올린 연주곡을 테스트 했을 때, 바이올린으로 인식된 결과(V_V)와 기타(V_G) 또는 피아노(V_P)로 인식된 결과를 보여준다. 최저인식률은 Mixture 수가 '1'일 때 71.5% 이었고, 최고인식률은 Mixture 수가 '16'일 때 95.5% 이었다. 비교적 인식오류의 빈도수가 적었다. 전체적으로는 Mixture 개수 변화에 따른 인식률이 고르지 못하고 편차가 컸다. 최고 인식률을 비교해 보더라도 바이올린은 비교적 높은 인식률을 보였지만, 피아노와 기타 간의 인식 오류가 상당한 빈도수를 나타냈다.

오디오 신호로부터 밴드 별 에너지만을 가지고 실험한 결과인 (그림 5)와 달리, (그림 6)은 앞에서 설명한 바와 같이 (그림 3)(e)와 같은 피크 위치와 관련된 확률적 가중치를 구한 후, 이를 밴드 별 에너지 값과 곱하여 (그림 4)와 같은 결과를 얻고, 이렇게 각 밴드 별 가중치가 적용된 에너지를 특징벡터(feature vector)로 하여 이들로 훈련된 Gaussian Mixture Model을 이용하여 통계적으로 연주악기를 인식하는 방법을 실험한 결과이다. (그림 6)의 (a)는 기타 연주곡을 테스트 했을 때, 최저인식률은 Mixture 수가 '1'일 때 67.5% 이었고, 최고인식률은 Mixture 수가 '13'일 때 94% 이었다. 여기서도 역시 대부분 바이올린보다는 피아노로 인식되는 오류의 빈도수가 많았다. (그림 6)의 (b)는 피아노 연주곡을 테스트 했을 때, 최저인식률은 Mixture 수가 '1'일 때 90.5% 이었고, 최고인식률은 Mixture 수가 '12' 또는 '13'일 때 99% 이었다. (그림 6)의 (c)는 바이올린 연주곡을 테스트 했을 때, 최저인식률은 Mixture 수가 '1'일 때 87.5% 이었고, 최고인식률은 Mixture 수가 '12' 또는 '13'일 때 99.5% 이었다. 전체적으로 Mixture 개수 변화에 따른 인식률이 악기에 관계없이 '1'에서 최저였고, '12'또는 '13'에서 최고를 나타냈고, 인식 성공률도 전체적으로 향상되었다. 특히 기타와 피아노의 구분 성공률이 최고 14%의 상대적 향상(G_G 최고인식률 참조)을 보였다.

위의 실험 결과로 볼 때, 기타와 피아노는 피크를 중심으로 그 주변의 신호패턴을 분석하는 것이 두 악기를 분류하는데 도움이 되었다. 즉 밴드 별 에너지라는 기본적인 특성



(그림 6) 피크 기반 확률적 가중치 적용 방법의 연주악기 인식률: 기타(G), 피아노(P), 바이올린(V)

만을 이용하더라도 피크와 관계없이 적용하는 것보다 피크를 고려해서 적용하는 것이 인식률 향상을 만들어 냈다. 이는 두 악기 인식에 있어서 피크 탐색의 중요성을 보여주는 것이다. 이와 같은 결과를 밴드 별 에너지 이외의 오디오 신호 특성들에 대해서도 적용하여 피크의 존재 여부를 기반으로 악기인식을 할 경우 어떤 향상을 보여줄 수 있는지 연구해 볼 흥미로운 주제가 될 것이다.

6. 결 론

본 논문에서는 UCC의 음원을 분류하는데 있어서, 사용자가 직접 연주하여 제작하는 콘텐츠에 가장 많이 사용되는 악기인 기타와 피아노, 그리고 바이올린의 독주 연주음악 오디오 음원을 이용하여 연주악기를 구분하는 연구의 결과를 보여주었다. 세 가지 악기는 도구와 방법은 다르지만 모두 현을 진동시켜서 소리를 발생시키고, 특히 기타와 피아노는 매우 흡사한 오디오 신호 패턴을 보였다. 두 악기의 연주로 만들어진 오디오신호를 분석한 결과 소리의 에너지 피크 지점을 중심으로 그 주변에서 가장 큰 에너지 변화 패턴의 차이를 보이는 것을 확인하였다. 이러한 점을 이용하기 위해 피크 위치에 대한 정보를 찾아내고 이를 활용한 결과 악기 인식률 향상을 확인하였다. 이번 연구결과를 토대로 기타와 피아노 연주 콘텐츠의 분류와 이와 관련된 UCC 메타데이터 자동생성에 큰 역할을 할 수 있다.

참 고 문 헌

[1] K. Martin and Y. Kim, "Musical Instrument Identification: A Pattern Recognition Approach," In Proc. of 136th Meeting of the Acoustical Society of America, 1998.

[2] A. Eronen and A. Klapuri, "Musical Instrument Recognition using Cepstral Coefficients and Temporal Features," In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol.2, pp.753-756, June, 2000.

[3] H. Hacihabiboglu and N. Canagarajah, "Musical Instrument Recognition with Wavelet Envelopes," In Proc. of EAA Convention, Forum Acusticum Sevilla, Sevilla, Spain, pp.16-20, Sep., 2002.

[4] S. Essid, G. Richard, and B. David, "Efficient Musical Instrument Recognition on Solo Performance Music using Basic Features," In Proc. of AES International Conference on Metadata for Audio, pp.2-5, June 2004.

[5] A. Livshin and X. Rodet, "Musical Instrument Identification in Continuous Recordings," In Proceedings of the 7th International Conference on Digital Audio Effects, 2004, pp. 222-226, October 2004.

[6] S. Essid, G. Richard, and B. David, "Musical Instrument Recognition by Pairwise Classification Strategies," IEEE Transactions on Audio, Speech, and Language Processing, Vol.14, No.4, pp.1401-1412, 2006.

[7] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "Instrogram: A New Musical Instrument Recognition Technique without using Onset Detection nor F0 Estimation," In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol.5, pp.229-232, 2006.

[8] A. Cont and S. Dubnov, "Realtime Multiple Pitch and Multiple Instrument Recognition for Music Signals using Sparse Non-Negative Constraints," In Proc. of International Conference on Digital Audio Effects, Bordeaux, France, DAFX.1-8, Sep., 2007.

[9] E. Hassan, "Instruments Recognition using Neural Networks and Spectral Information," In Proc. of International Conference on Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia, March 2007.

[10] Q.-Z. Wu, H.-Y. Cheng, and B.-S. Jeng, "Motion Detection via Change-Point Detection for Cumulative Histograms of Ratio Images," Pattern Recognition Letters, Vol.26, pp.555-563, 2005.

[11] R. Duda, D. Stork, and P. Hart, "Pattern Classification," Wiley-Interscience 2/E, 2000.

[12] 권순일, 박지형, 박능수, "한국어 발화음성에서 중점단어 탐색을 위한 기본주파수에 대한 연구," 정보처리학회논문지B, Vol. 15, P.595-602, 2008



권 순 일

e-mail : sikwon@sejong.ac.kr

1998년 연세대학교 전자공학과 졸업(학사)

2000년 미국 University of Southern California 전기공학과 졸업(공학석사)

2005년 미국 University of Southern California 전기공학과 졸업(공학박사)

2005년~2006년 삼성전자 책임연구원

2006년~2009년 한국과학기술연구원(KIST) 선임연구원

2009년~현 재 세종대학교 디지털콘텐츠학과 교수

관심분야: 음성인식, 화자인식, 음성합성, 음성/오디오 인터페이스, HCI, HRI 등



박 완 주

e-mail : wanjoo@kist.re.kr

2006년 한국항공대학교 항공전자공학과
(학사)

2008년 서강대학교 전자공학과 (공학석사)

2008년~현 재 한국과학기술연구원(KIST)
연구원

관심분야: 머신러닝, 패턴인식, Haptics, HCI 등