

음악의 특성에 따른 피아노 솔로 음악으로부터의 멜로디 추출

최윤재⁰, 박종철

KAIST 전산학과

{yjchoi⁰, park}@nlp.kaist.ac.kr

Extracting Melodies from Piano Solo Music Based on Characteristics of Music

Yoonjae Choi⁰ Jong C. Park

Department of Computer Science, KAIST

인터넷의 발달로 멀티미디어 자료의 검색 및 활용 방법에 대한 연구가 활발히 진행되고 있다. 특히 디지털 음반 시장의 빠른 발전으로 인해 음악 검색 및 추천에 대한 수요가 계속해서 증가하고 있는데 이러한 서비스를 수행하는 음악 기반 응용 시스템의 성능 향상을 위해서는 일반적인 음악의 형태인 다음(Polyphonic)¹ 음악에서 멜로디²를 추출하는 과정이 필수적이다. 본 논문에서는 다음의 복잡도가 높고 넓은 음역을 가지는 음악을 만들 수 있는 피아노 솔로 음악에서 멜로디를 추출하는 방법을 제안한다.

음악으로부터 멜로디를 추출하는 방법은 크게 두 가지로 나눌 수 있다. MP3나 Wave 포맷의 파일로부터 오디오 정보를 받아들이고 신호처리 과정을 통해 멜로디를 추출하거나 미디(MIDI) 파일과 같이 음악을 상징적인 기호로 표현한 파일을 받아들이고서 멜로디를 추출할 수 있다. 오디오 시그널에서 멜로디를 추출하는 방법은 실제로 녹음된 음악에서 멜로디를 추출할 수 있다는 장점이 있지만 음악 자체의 특성을 효과적으로 사용하기 어렵다는 단점이 있다. 본 연구에서는 멜로디의 효율적인 추출을 위해서 음악의 특성을 우선적으로 고려하는 것이 중요하다고 보아 음표로 표현되어 있는 음악으로부터 멜로디를 추출하는 방법에 대해 논의하고자 한다. 특히 본 연구에서는 음악의 특성에 따라 음악을 세 가지 형태로 분류하여 각 유형에 맞는 원칙을 사용하여 멜로디를 추출하는 방식을 제안한다.

본 논문은 피아노 솔로 음악의 구조적 특징을 고려해서 음악을 다음과 같이 세 가지 유형으로 나누는 방식을 제안한다.

1. 동시에 발생하는 음들의 지속 시간이 모두 동일하고 그 외에 음들 간의 겹침이 없는 경우 (Equal Occurrence: E 유형)

2. 멜로디가 상단부에 위치하는 경우 (High Melody: H 유형)

3. 멜로디가 하단부에 위치하는 경우 (Low Melody: L 유형)

E 유형의 음악은 반주와 멜로디가 확연히 구분되지 않고 전체가 화음으로만 구성된 구조다. 동시에 시작된 음들은 대개 동시에 끝나고 다음에 발생하는 음들과는 겹침이 없기 때문에 음들 간의 수직적인 겹침은 존재해도 수평적인 겹침은 존재하지 않는다. 이와 같은 특성을 이용해 입력된 음악이 E 유형인지 판단하고, E 유형의 음악임이 확인될 경우 E 유형의 멜로디 추출 알고리즘을 적용한다. E 유형의 멜로디 추출 알고리즘으로는 기존의 멜로디 추출 알고리즘인 Skyline Algorithm을 사용하기로 한다.

H 유형의 음악은 멜로디가 상단에 존재하는 음악으로 대부분의 피아노 음악은 H 유형에 속한다고 볼 수 있다. 우리가 일상적으로 접하는 음악의 대부분이 H 유형에 속하는 만큼 H 유형의 음악에는 여러 가지 패턴이 존재하기 때문에 E 유형의 경우와 같이 공통적인 특성을 발견하기 어렵다. 그러므로 H 유형의 판별 모듈은 따로 제작하지 않고 E 유형과 L 유형으로 판단되지 않은 나머지 음악을 H 유형에 속한다고 판단한다. H 유형의 음악으로부터 멜로디를 추출하는 경우에는 Skyline Algorithm을 활용하되 특별한 경우가 아니라면 음의 지속 시간을 지켜야 한다.

¹ 다음(Polyphonic) 음악은 두 가지 이상의 음들이 동시에 발생하는 음악을 뜻한다. 이와 달리 단음(Monophonic) 음악은 한 번에 두 개 이상의 음들이 발생하지 않는 음악이다. 예를 들어 한 사람의 노래는 단음 음악에 해당한다.

² 여기서 멜로디라 함은 하나의 개체로 인식되는 연속된 음들을 뜻한다. 화성이 복수음의 수직적 결합이라면 멜로디는 음들의 수평적, 단음적 배열이다.

L 유형의 음악은 멜로디가 하단에 존재하는 경우로서 흔히 볼 수 있는 음악의 구조는 아니다. 상단의 음들이 일정하고 하단의 음들이 더 흥미로운 패턴을 이루고 있을 때 하단의 음들이 멜로디로 인식된다. 상단의 음들이 일정하다는 것은 곧 상단의 반주음들의 지속 시간이 일정하다는 뜻이다. 상단부의 음들이 반주를 구성하고 하단부의 음들이 멜로디를 구성할 때에는 상단부의 반주음들의 지속 시간이 전부는 아닐지라도 매우 일정하다. 또 많은 경우 반주음들의 진행 속도가 멜로디음들의 진행 속도보다 빠르다. 이와 같은 특성을 이용해 입력된 음악이 L 유형의 음악인지 판단할 수 있다. 입력된 음악이 L 유형임이 확인된다면 멜로디 추출 알고리즘을 적용한다. 멜로디 추출 시에는 반주에 해당하는 음들을 인식한 후에 반주음들을 제거하고 남은 음들 중에서 음높이가 가장 높은 음을 채택한다.

본 연구에서 구현한 시스템의 구조는 그림 1과 같다. 시스템은 리눅스를 운영체제로 하고 파이썬을 이용해서 구현했다. 전처리 과정은 미디 파일을 입력 받아서 처리하고 유형 판별 모듈은 입력 받은 음악이 E, H, L 유형 중에서 어디에 속하는지 판단한다. 멜로디 추출 모듈은 판단된 음악의 유형에 따라 멜로디 추출 알고리즘을 적용하여 멜로디를 추출한다. 미디 파일 생성 모듈은 추출된 멜로디를 다시 사용자가 들을 수 있는 미디 파일 형태로 제작해주는 역할을 한다. 그림 2는 리스트의 사랑의 꿈을 미디 파일 형태로 입력 받아 멜로디 추출 과정을 거친 후의 결과 화면이다.

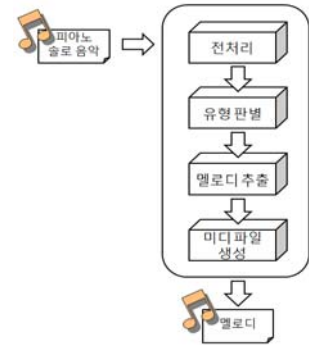


그림 1. 시스템 구조도

시스템의 평가는 유형 자동 판별 모듈과 멜로디 추출 모듈의 성능을 각각 측정하는 방식으로 한다. 테스트 대상은 연구 과정에서 사용한 음악 22개로 구성된 Training Set과 연구 과정에서 사용하지 않은 음악 20개로 구성된 Test Set으로 구성된다.

그림 3은 유형 자동 판별 모듈의 실험 결과를 나타낸다. 실험은 음악의 마디를 단위로 N-Gram 형식과 같이 보는 대상을 한 마디씩 증가시키면서 진행되었다. 두 그룹 모두 마디를 많이 볼수록 유형 판별의 정확도가 증가하며 로그 형태의 곡선을 보이고 있다.

```

Average_samp_duration_ratio: 0.052941
Average_overlap_ratio: 0.723691
Average_pitch_difference: 11.000000
Standard_deviation: 1.793051
L_type
[51, 0, 190, 110, 190, 60, 192, 766, 110, 574, 60, 768, 1342, 110, 574, 60, 1344, 2302, 110, 958
, 60, 2304, 2494, 110, 190, 60, 2496, 2878, 110, 382, 60, 2880, 3070, 110, 190, 62, 3072, 3454
, 110, 382, 60, 3456, 3646, 110, 190, 60, 3648, 4232, 110, 574, 53, 4224, 4414, 110, 190, 60, 4
416, 4512, 110, 96, 72, 4512, 4608, 110, 96, 53, 4608, 4798, 110, 190, 53, 4800, 4990, 110, 190
, 55, 4992, 5182, 110, 190, 56, 5184, 5374, 110, 190, 60, 5376, 5758, 110, 382, 56, 5760, 5950
, 110, 190, 56, 5952, 6334, 110, 574, 32, 6328, 6718, 110, 190, 60, 6720, 6912, 110, 96, 69, 68
16, 6912, 110, 96, 51, 6912, 7102, 110, 190, 60, 7104, 7278, 110, 574, 60, 7280, 8254, 110, 574
, 60, 8256, 9214, 110, 958, 60, 9216, 9406, 110, 190, 60, 9408, 9790, 110, 382, 60, 9792, 9982
, 110, 190, 61, 9984, 10366, 110, 382, 60, 10368, 10558, 110, 190, 61, 10560, 11134, 110, 574
, 34, 11136, 11326, 110, 190, 59, 11328, 11424, 110, 96, 65, 11424, 11520, 110, 96, 53, 11520
, 11710, 110, 190, 53, 11712, 11904, 110, 192, 85, 11904, 12096, 110, 192, 56, 12096, 12288, 11
0, 190, 60, 12288, 12672, 110, 384, 58, 12672, 12862, 110, 190, 56, 12864, 13054, 110, 190, 6
3, 13056, 13152, 110, 96, 60, 13152, 13148, 110, 96, 63, 13248, 13344, 110, 96, 60, 13344, 134
30, 110, 96, 32, 13430, 13620, 110, 190, 60, 13620, 13728, 110, 96, 63, 13728, 13824, 110, 96
, 60, 13824, 13920, 110, 96, 56, 13920, 14014, 110, 94]
    
```

그림 3. 리스트의 사랑의 꿈의 멜로디 추출 결과

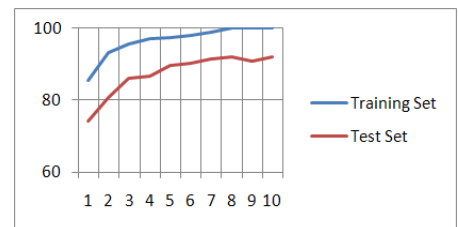


그림 2. 유형 판별 모듈의 실험 결과

멜로디 추출 모듈 실험에서 Training Set은 재현율(Recall) 99.05%, 정확도(Precision) 88.62%를, Test Set은 재현율 94.76%, 정확도 89.93%를 기록했다. 전체적으로 높은 수치를 기록했지만 예외적인 경우도 있었는데 쇼팽의 혁명 연습곡과 같이 매우 빠른 반주음들과 휴식 기간이 잦은 멜로디를 갖는 음악들의 경우에는 정확도가 50% 전후를 기록하기도 했다. 또 애초에 유형 분류가 잘못된 음악의 경우에는 재현율과 정확도 모두 20% 미만을 기록하기도 했다. 잘못된 유형 분류는 왼손 반주의 진행이 느리고 오른손 멜로디음들의 지속 시간이 일정하고 진행이 빠른 H 유형의 음악을 L 유형의 음악으로 잘못 판단했기 때문에 발생했다. 이와 같은 유형 분류의 오류는 추후 연구를 통하여 해결되어야 할 것이다.

본 논문에서는 피아노 솔로 음악을 음악의 특성에 따라 세 유형으로 나누어 분석하여 유형별 특징을 조사하고 각 유형으로부터 멜로디를 추출하는 방법에 대해 논의하였다. 추후에는 하나의 음악 내에 둘 이상의 유형이 혼재하는 경우를 다룰 수 있도록 음악의 의미 있게 분할하는 연구를 진행할 계획이다. 그리고 유형 판별이 아직 완벽하게 이루어지는 것이 아니므로 더욱 많은 피아노 음악을 분석해 유형 판별 모듈의 성능을 향상시키면 시스템의 전체적인 완성도가 한층 높아질 것이라고 기대한다.

감사의 글

논문 작성에 도움을 주신 이호준님에게 감사드립니다. 본 연구는 21세기 프론티어 연구개발 사업(인간 기능 생활지원 지능로봇 기술개발사업)을 통한 지식경제부의 지원을 받아 수행되었습니다.