리듬기반 인터액티브 음악 플레이어를 위한 음표 위치 추적 알고리즘

Note Tracking and Localization Algorithm for Interactive Rhythm-based Music Player

김재홍 Jaehong Kim*, 박재성, Jaesung Park*, 이준성, Junseong Lee*, 차동훈, Donghoon Cha*
서울과학고등학교

김정현, G. Jounghyun Kim**
고려대학교 디지털체험 연구실

요약 기존의 음악 플레이어들은 녹음 되거나 디지털적으로 캡쳐 된 음악정보를 재생 하여 사용자들이 "한 방향" 으로의 수동적인 감상을 가능하게 하였다. 본 논문에서는 mp3 나 wav 로 형태로 실제로 연주 되고 녹음 된 음악에서의 특정 음표의 시간적인 위치를 찾아내는 알고리즘을 소개 하도록 한다. 찾아내고자 하는 음표의 악보를 이용하면, 우선 주어진 녹음 된 음악 파일에서의 해당 음표의 위치를 시간적인 순서대로 예측 할 수 있다. 그러나, 연주/녹음 된 음악은 악보에 나와 있는 대로 연주 되지 않고 대부분 시간적으로 혹은 심지어 내용적으로 변화가 있게 마련이다. 따라서 추가적인 분석을 통하여 음표의 정확한 위치를 찾아나가게 되고, 그 위치로부터 이러한 예측 및 교정작업 (prediction/correction)을 계속적으로 수행하게 된다. 이러한 부가적인 정보를 이용하여 사용자가 음표의 위치에 (즉 리듬에) 기반한 인터액션을 통하여 실제 음악을 연주하는 듯 한 사용자 경험을 줄 수 있다.

Abstract Conventional music players offer simple replay and one way entertainment. The paper presents an algorithm to extract, within a digitally recorded music file, the temporal information of a sequence of target notes (i.e. melody). We assume to have the score (e.g. MIDI or printed score), and using this information, it becomes possible to first sequentially predict the probable location of the target notes. However, recorded music is hardly performed according to the score, especially temporally. Thus, additional analysis is carried out to hone in on the exact location of the target note from the initially predicted location. This prediction and correction process is repeated to find one note after another. This allows us to develop an interactive music player that is enacted by rhythmic interaction, and induce a new user experience, i.e. as if one is playing the music oneself.

핵심어: Keywords *Interactive Music Player, Pitch Detection, Note Tracking, Bayesian Filter*

본 논문 2008 년 과학재단 지원 과학고 영재교육 (R&E) 내실화 프로그램 지원에 의하여 연구되었음

*주 저자: 서울과학고등학교 김재홍 (ekzm0204@naver.com), 박재성 (pjsdream2001@hanmail.net), 이준성(com0021@gmail.com), 차동훈 (danny831@empal.com)

**교신 저자: 고려대학교 정보통신대학 김정현 교수 (gjkim@kora.ac.kr)

1. 서론

기존의 음악 플레이어들은 녹음 되거나 디지털적으로 캡쳐 된 음악정보를 재생 하여 사용자들이 "한 방향" 으로의 수동적인 감상을 가능하게 하며, 이는 에디슨의 전축의 발명 이후, 인간의 음악 감상의 주 형태를 이루어 왔다. 물론 가장 이상적인 형태의 음악 "감상" 은 직접 연주를 통한 "양방향" 감상이 되겠으나, 이는 사용자로 하여금 많은 연습과 비용을 (악기, 장소, 인원, 시간) 들게 한다. 이를 개선하기 위하여 리듬을 이용한 BeatMania [1]와 같은 게임이나 Fakeplay [2] 와 같은 리듬 기반의 미디 플레이어들이 소개 되었다. 즉 사용자가 음악의 리듬에 익숙 한 경우, 리듬에 맞춘 인터액션을 통하여 음악 플레이어를 구동하여 간접 연주의 경험감을 주는 접근 방법이라고 할 수 있다. BeatMania (및 DDR 과 같은 BeatMania 의 아류들) 의 경우, 이들이 상용 게임이므로 리듬 인터액션에 필요한 정보와 재생하고자 하는 음악과의 관계 정보를 개발자가 직접 수동 만들어 내어 제공 하여야 하고, 따라서, 사용자는 개발자가 제공하는 음악만을 해당 게임 환경에서만 재생 할 수 있다. Fakeplay 의 경우 미디를 이용하기 때문에 범용성은 뛰어난 반면 미디 수준의 인공적인 음질 때문에 각광을 받지 못하고 있다.

이 논문에서는 미디 기반의 Fakeplay 를 mp3 나 wav 로 형태로 디지털적으로 녹음 된 음악 기반으로 확장하기 위하여, 녹음 된 음악 파일에서 리듬에 해당 하는 노트의 위치를 추출 해 내는 알고리즘을 소개 하도록 한다. 이러한 알고리즘을 사용하여 기존의 mp3 플레이어는 리듬과 기타 Expression 을 위한 인터액션을 통하여 음악을 "플레이" 하여, 실제 음악을 연주하는 듯한 사용자 경험을 줄 수 있을 것이다 (그림 1).

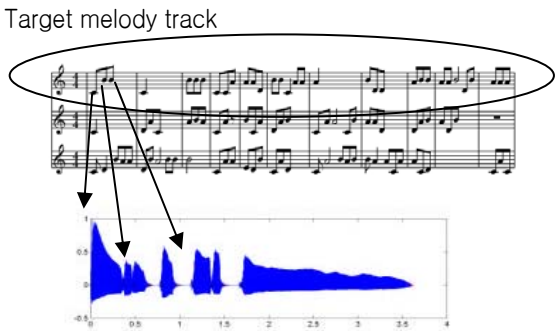


그림 1: 녹음 된 음악으로부터의 멜로디 리듬 음표의 Localization 문제. 이를 기반으로 실제 연주를 녹음한 음악도 리듬에 맞추어 Interactive 하게 플레이 할 수 있다.

2. 관련연구

본 연구는 앞서 언급 된 미디 기반의 Fakeplay 를 확장 하기 위해서 수행 되었다. 미디 기반의 Fakeplay 는 친숙한 음악의 멜로디 리듬에 맞추어 사용자가 인터액션을 취하면 (예: 컴퓨터 키보드, 압력 센서, 미디 건반, 미디 관악기등) 주어진 음악의 미디 파일이 플레이 되는 시스템이다 [2]. Fakeplay 는 좀 더 앞서 Konami 에서 출시 된 BeatMania 와 비슷하다. BeatMania ,게임에서는 음악이 플레이 되면서 리듬 정보에 맞추어 사용자가 적절한 반응을 하면 점수를 얻는 방식이며, 인터액션을 취하지 않아도 음악이 계속 연주 되기 때문에 음악을 직접 연주 한다는 느낌은 덜 받는다[1]. Raphael 은 "Music Plus One" 이라는 시스템은 실시간에 연주 되고 있는 모노 (단일악기, 예: 오보에) 기반의 음악을 Dynamic Bayesian Network 를 이용하여 음악의 미세한 템포와 감정 변화를 인식하여 녹음 된 반주 음악 파일을 같은 템포 및 감정으로 같이 연주 해 주는 시스템인데, 협주 연습에 매우 좋은 역할을 할 수 있다 [3]. Raphael 이 사용한 Dynamic Bayesian Network 는 이 논문에서 채택한 Bayesian Filter 방법과 같다고 할 수 있다. 이 논문에서는 음표의 위치 예측 이후 좀 더 정확한 위치의 파악을 위해 추가적인 서치를 하게 되는데, 이때 사용하는 음표의 매칭 특성으로서는 주파수 분석을 통한 Fundamental 주파수의 에너지 분포, Autocorrelation 분석을 통한 주기성 (Periodicity) 및 ADSR (Attack-Decay-Sustain-Release) 패턴등을 흔히 쓰게 된다 [4].

3. 알고리즘

본 논문에서 제안하는 알고리즘은 Bayes Filter 에 기반한 것으로 그림 2 에 나타난 것과 같이 Prediction-Correction 형태를 띠고 있다. 즉 음표의 위치를 나타내는 x(t)가 관찰 (분석) 된 음악의 결과에 따라 갖는 확률이 최대가 되는 시간적 지점을 계산 하는 것이다.

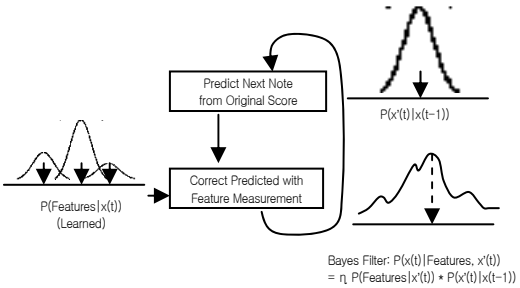


그림 2: Bayes Filter 를 이용한 리듬 음표 Localization 알고리즘.

3.1 예측(Prediction)

제안하는 알고리즘은 우선 실제 악보와 녹음 된 음악의 간단한 시간 분석을 통하여, 녹음 된 음악에서의 시작점/끝점 및 1 박자 당 평균 시간을 추출 하는 것으로 시작 한다 (Seconds per Beat). 이 값은 각각의 음표의 시간적 정보가 추출 되면서 약간씩 변화 할 수도 있다. 이를 바탕으로, 실제 녹음 된 음악에서의 최초 리듬 음표의 시작점으로부터 다음 해당 리듬 음표의 위치를 순차적으로 음악이 끝날 때까지 "예측" (Prediction) 할 수 있다. 그러나 실제 연주 된 음악에서는 미디와 달리 많은 시간적, 심지어 내용적 "약간의" 변화가 있기 마련이다 (예: 템포나 강약의 변화). 따라서 이 예측 데이터만을 가지고는 정확한 리듬 음표의 시간적 위치를 알아 낼 수 없다. 그림 2 에서의 보이는 것과 같이 예측 된 음표의 위치의 확률 분포는, P(x' (t)|x(t-1)), Gaussian 분포를 사용한다.

3.2 교정 (Correction)

그래서, 예측 된 위치 주위에서 (해당 음표의 길이와 비슷한 크기의 서치 윈도우를 염) 해당 리듬 음표와 가장 비슷한 음표를 여러 가지 특성에 기반 하여 찾게 된다. 크게 사용하는 두 가지의 매칭 기준은 주파수 분석 결과와 ADSR 패턴이다.

주파수 분석의 경우, 해당 리듬 음표와 동시에 연주 되는 반주 부분의 소리 때문에 단순한 Monophonic pitch 매칭 혹은 분석은 가능 하지 않다. Polyphonic 사운드의 (혹은 화음)에서의 주파수 Signature 는 알려지고 있지 않다. 따라서, 본 논문에서는 기계학습 혹은 Data Base 접근 방법을 사용한다. 즉, 미디 파일을 wav 파일로 변환하여 이 파일이 악보와 똑같이 연주 되었다는 가정하에, 해당 리듬 음표의 주파수 Signature Data Base 를 미리 저장하는 것이다. 좀 더 구체적으로는 가장 큰 에너지를 갖는 주파수 10 개를 저장 한다. 이는 곧 P(음표 Feature | x(t)) 의 확률 분포를 제공 한다. 목표 분석구간의 가장 큰 주파수 10 개와 비교하여 확률 분포를 결정한다.

그러나, 실험 결과, 주파수별 에너지 분포는 음표의 위치를 정확하게 발견하는데 크게 도움이 되지 않았다. 그림 3 에서 두 구간이 보여 지고 있는데, 왼쪽의 구간은 목표 음표의 구간과 약간 겹쳐 있고, 오르쪽의 구간은 거의 완전히 겹쳐 있다. 그러나 두 구간의 주파수 분석 결과는 비슷 하였다.

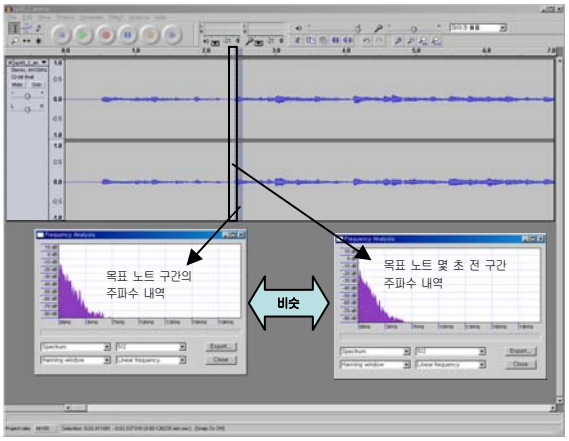


그림 3: Polyphonic 사운드 주파수 분석의 어려움.

ADSR 분석은 예측 된 음표 위치 주위, 서치 윈도우 안에서 최대의 음파량을 갖는 위치를 찾는 것이다. Periodic 한 음파의 절대값을 구하고, 이를 Low Pass 필터링을 통하여 부드러운 음파 프로파일을 얻는다. 그 다음 이를 미분하고 Zero Crossing 을 찾은 다음, 그 중 최대의 음파량을 갖는 음파 Peak 를 찾는다. 주파수 분석이 별로 효용이 있지 않아 주파수와 ADSR 사이에서 후자에 더 많은 가중치를 (1:9) 주고 P(음표 Feature | x(t)) 의 값을 구한다. 그러나 어려운 점은 Noise 때문에 필터링을 했을 때 중요한 Peak 가 사라질 수도 있어서 Ad Hoc 하게 필터링의 정도를 결정 해야 한다는 점과 Ad Hoc 하게 결정해도 구간 별로 성능의 차이가 많이 날 수 있다는 것이다 (그림 4 참조).

그림 4: ADSR Peak 찾기.

4. 실험

3 장에 소개 된 알고리즘은 Matlab 로 구현 되었으며, Polyphonic 한 (최대 4 개 음의 화음) 단일 악기 (피아노) 음악으로 (베에토벤 피아노 소나타 작품 번호 Op. 49, 제 2 악장) 테스팅 되었다. 그림 5 는 2 개의 각기 다른 wav 파일에 (다른 사람에 의해 연주 됨) 대하여 알고리즘의 성능을 보여 주고 있다. 첫 번째 연주 파일에서 두 번째 연주 파일에 비해 템포의 변화가 대체로 많지 않아 더 좋은 정확도를 보여 주고 있다. 즉 두 번째 연주가의 경우, 초기부터 평균 속도에 비하여 많이 느리게 연주 하면서, 예측이 잘못 되고, 따라서 초기 부터 트래킹이 잘 되지 않고 있다. 이에 비하여 첫 번째 연주에서는 트래킹이 어느 정도 잘 되지만, 이 경우도 시간이 흐름에 따라, 약간씩 부정확하게 찾아진 위치에 의하여 그 다음 위치가 영향을 받아 점점 트래킹이 어긋나는 결과를 얻었다.

5. 결론

음악에서의 트래킹은 공간과 달리 매우 정확한 트래킹을 요구 한다. 이는 인간의 청각의 특성상 Pitch, Timing 등의 Sensitivity 가 매우 높기 때문이다. 녹음 된 음악에서의 화음 트래킹은 우서 주파수 분석에 거의 불가능 한 점과 ADSR 패턴 분석도 Noise 에 의해 어렵고, 또한 음악연주의 순차적 특성상, 시간적으로 후의 트래킹이 전 트래킹에 의존적일 수 밖에 없어 많은 어려움을 갖게 된다. 앞으로, 트래킹의 성능을 개선하고, 인터액티브 mp3 player 를 개발 하기 위해서는, 음표를 구분 할 수 있는 또 다른 특성을 계속 찾고, 동시에 자동 알고리즘 보다는, 사용자로부터 최소한의 도움을 얻는 Semi-automatic 알고리즘 및 Interaction Editor 를 개발 하려고 한다.

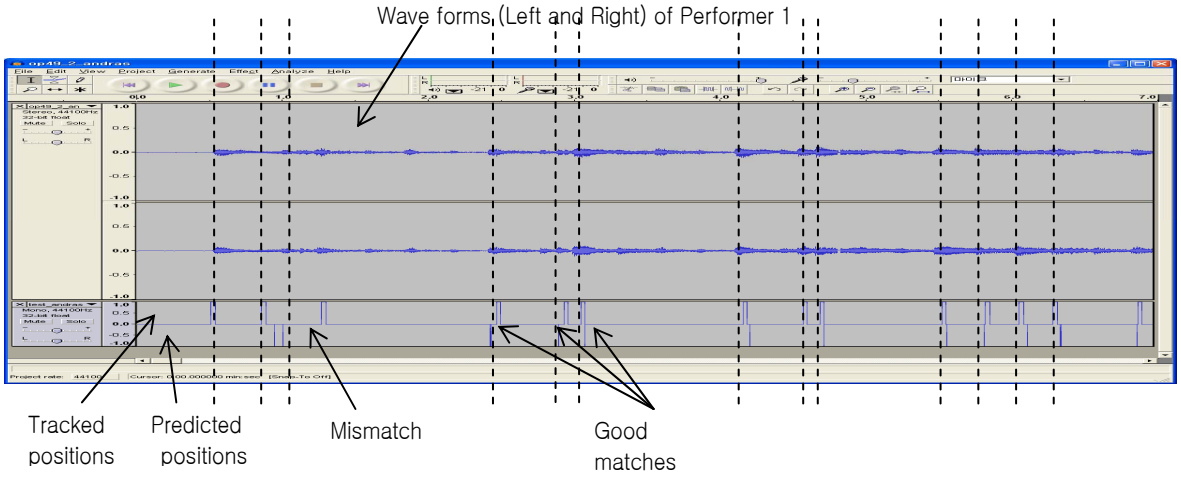
참고문헌

[1] Konami, "Beatmania", www.konami.com

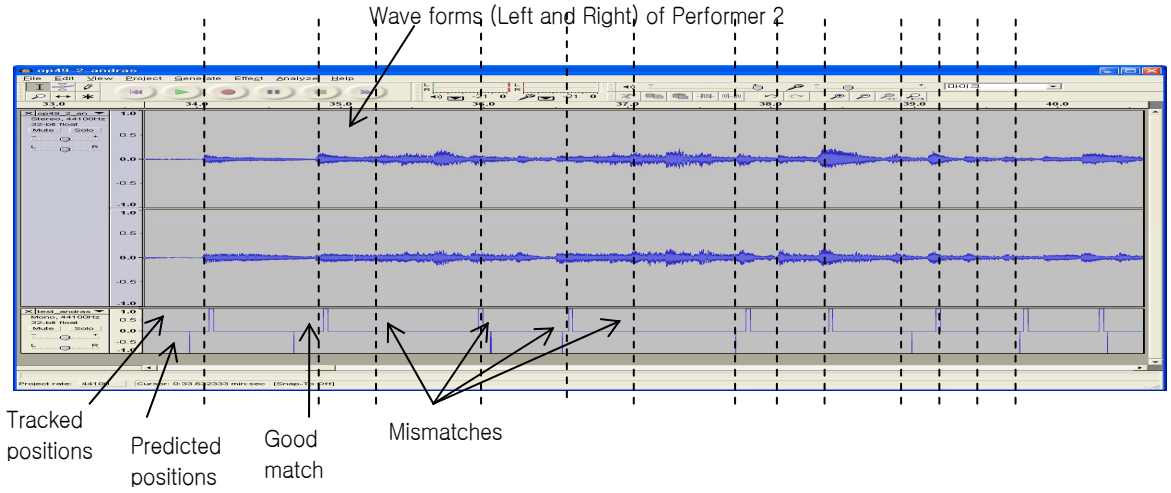
[2] G. J. Kim, "Active Music Appreciation with Fakeplay," Virtual System and Multimedia 2002.

[3] C. Raphael, "Music Plus One: A System for Flexible and Expressive Musical Accompaniment," Proc. of the Intl. Computer Music Conf., Havana, Cuba, 2001.

[4] C. Roads, "The Computer Music Tutorial," MIT Press,1996.



(a) 연주가 1



(a) 연주가 2

그림 5: 알고리즘의 성능 (a) 연주가 1 (템포 변화 적음), (b) 연주가 2 (템포 변화 많음). 두 경우 모두 스테레오 음파 (Left/Right Channels)을 보여주고 있고 각 음파 밑의 사각 음파는 트래킹 결과로 만들어진 인공 음파이다. 위쪽은 찾아진 음표의 위치를 표시 하며, 아래쪽은 예측된 위치를 표시 한다 (예측 위치에 해당 하는 사각 음파가 트래킹 결과 사각 음파에 의해 덮어 씌어져 안 보이기도 한다). 점선은 음표가 있어야 하는 위치를 뜻한다.