

내추럴 유저 인터페이스를 이용한 제스처 기반 작곡 시스템의 구현

Implementation of Music Composition System Based on Gesture Recognition Using Natural User Interface

윤겨레

Gyeore Yun

포항공과대학교 컴퓨터공학과
Dept. of Computer Science and
Engineering, POSTECH
ykre0827@postech.ac.kr

유용재

Yongjae Yoo

포항공과대학교 컴퓨터공학과
Dept. of Computer Science and
Engineering, POSTECH
dreamseed@postech.ac.kr

최승문

Seungmoon Choi

포항공과대학교 컴퓨터공학과
Dept. of Computer Science and
Engineering, POSTECH
choism@postech.ac.kr

요약문

본 연구에서는 키넥트를 사용해, 전문적인 지식이 없는 비전문가도 두 팔과 손의 제스처를 이용해 음악을 손쉽게 작곡할 수 있는 내추럴 유저 인터페이스(NUI; Natural User Interface) 기반의 시스템을 구현하였다. 키넥트 센서로부터 얻어진 데이터를 은닉 마르코프 모델 (Hidden Markov Model)을 이용하여 모델링 및 학습함으로써 사용자의 동작을 인식하고, 각 동작에 음악의 구성요소(멜로디, 리듬, 음색, 화성 진행)를 대응시켰다. 이들 중에서, 화성 진행의 대응을 위하여 으뜸음을 기준으로 하는 도수(Degree) 화음의 개념을 도입, 이를 이용하여 대중음악의 화성 진행과 반복과 같은 음악의 특성을 활용하여 마르코프 연쇄 (Markov Chain)를 모델링하였다.

주제어

인간 컴퓨터 상호작용(Human Computer Interface), 내추럴 유저 인터페이스(Natural User Interface), 작곡 시스템(Music Composition System)

1. 서론

음악은 박자, 선율, 화성, 음색 등을 일정한 형식으로 나타내어 사상, 감정을 나타내는 종합 예술 장르로서, 인간의 정서적 가치를 충족시키는 데 기여하였다. 하지만 작곡은 오랫동안 전문 작곡가에 의존하여 왔으며, 현재까지도 음악적 지식을 가지고 있는 전문가의 영역으로 생각하는 경향이 남아있다 [1].

그러나, 최근 들어 컴퓨터 음악 인터페이스와 가상 악기의 발전으로 비전문가가 작곡을 공부하여 음악을 창작 또는 재구성하는 시도가 점차 늘고 있다. 또한 지식이 없는 일반인의 음악 창작에 대한 욕구를 만족시키기 위하여, 최소한의 정보를 이용해 음악을 자동으로 생성시켜주는 연구도 진행되었다. 이윤재 등은 허밍을 이용해 악보를 추정하여 음악을

자동으로 생성해주고, 박자에 어긋나거나 음악적 화성에 어울리지 않는 멜로디를 자동으로 보정하는 알고리즘을 제안하였다 [2]. 또, Simon 등은 사용자의 보컬 멜로디에 어울리는 코드를 자동으로 선택해주는 시스템을 제안하였다 [3].

게다가, 인간-컴퓨터 상호작용(HCI)에 관한 연구가 최근 여러 분야에서 활발히 진행되면서, 센싱 장치(Sensing Device)를 이용한 행동 기반의 연주와 작곡에 대한 연구가 주목을 받고 있다. Winkler 등은 사람의 제스처를 어떻게 음악적 요소에 대응시켜 작곡을 할 것인지에 대한 연구를 수행하였으며 [4], Hsu 등은 키넥트(Kinect) [5]를 이용한 제스처 인식(Gesture Recognition)을 통해 드럼, 기타 등의 악기를 연주할 수 있는 시스템을 제안하였다 [6].

그러나 이러한 연구들은 단순히 악기 연주를 대체하는 수단으로써 주로 활용되고 있기 때문에 연주법 등 악기에 대한 기본적인 지식이 없이 사용하기에는 여전히 어려움이 있으며, 작곡 시스템으로서의 효용은 부족하다. 본 연구에서는 이러한 장벽을 최소화하고, 앞서 언급한 음악의 창작에 대한 욕구를 충족시키기 위하여, 음악적 지식이 없이도 작곡을 할 수 있도록 내추럴 유저 인터페이스(NUI; Natural User Interface)를 사용하는 작곡 시스템을 개발하였다. 대중음악의 화성 진행을 마르코프 연쇄(Markov Chain)로 모델링하고, 자동으로 자연스러운 화성 진행을 생성하며 그에 어울리는 멜로디를 선택해주는 알고리즘을 개발하였다. 이를 이용해 제스처 기반의 작곡 프로그램을 구현하였다. 키넥트 센서로부터 획득한 관절 정보(Skeleton Data)와 은닉 마르코프 모델(HMM; Hidden Markov Model)을 이용해 사용자의 동작을 인식하며, 특정 동작을 수행하였을 경우에 음악의 구성요소 [7]인 멜로디, 리듬, 음색, 화성 진행을 생성하고 이를 소리로 출력하였다.

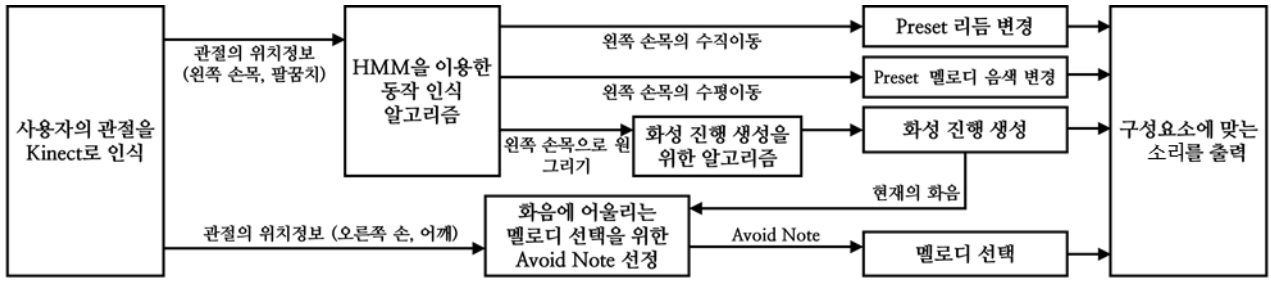


그림 1. 본 시스템 구성의 블록 다이어그램

2. 시스템의 구성 및 핵심 알고리즘

본 시스템 구성의 블록 다이어그램은 그림 1 과 같다. 우선, 키넥트를 이용하여 사람의 양 팔 손목과 왼팔의 팔꿈치, 오른팔의 손과 어깨의 위치 정보를 획득한다.

HMM 은 왼쪽 손목과 팔꿈치의 위치를 이용해 3 가지 동작을 인식(수직 이동, 수평 이동, 원 운동)하며, 이 동작들은 리듬, 음색, 화성 진행의 변경에 각각 대응된다. 리듬과 음색의 경우는 시스템 상에 미리 저장한 리듬들을 Switch 하며, 화성 진행의 경우 후술할 화성 진행 생성 알고리즘을 이용한다. 오른손의 높낮이는 마지막 음악의 구성요소인 멜로디를 생성하는 데 사용되며, 이 중 현재 화성과 어울리지 않는 음은 제외된다. 최종적으로 각 구성요소에 해당하는 소리를 출력, 하나의 완성된 음악을 출력한다. 각 동작에 대한 정보는 표 1 과 같다.

2.1 HMM 을 이용한 동작 인식 알고리즘

HMM [8]은 패턴이나 동작을 확률적으로 추론하고 이를 바탕으로 자동으로 학습하는데 적합한 모델로, 본 연구에서 사용자가 왼팔로 원을 그리거나 수직 및 수평으로 움직이는 동작을 인식하기 위해 사용되었다. 본 연구에서 사용한 HMM 의 모델링에서는 표 1 과 같이, 각각 동작에 따라 State 의 수와 관측열의 길이, 그리고 심볼의 수를 정하였다. 예를 들어, 그림 2 와 같이 왼쪽 손목-팔꿈치 구간으로 원을 그리는 동작 (1)의 경우, State 를 8 개로 나누고 초기 State Transition 을 설정하였다.

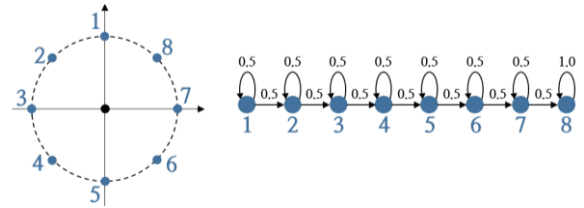
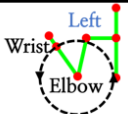
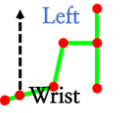
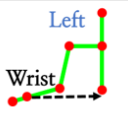




그림 2. 동작 (1)의 State 와 State Transition

표 1. 본 연구에서 사용하는 동작과 해당 동작에 대한 정보

동작	신체 부위	제스처	구성 요소	결과	HMM 모델링을 위한 정보	State 개수	Symbol 개수	관측열 길이
(1)	왼쪽 손목-팔꿈치	 시계 반대방향의 원을 한바퀴 그린다.	화성 진행	화성을 진행을 생성하고, 음악을 재생한다.	손목의 팔꿈치에 대한 위치	8	10	10
(2)	왼쪽 손목	 수직 위로 움직인다.	리듬	Preset 리듬을 바꾸거나 음악을 정지한다.	이전 프레임에 대한 손목의 방향	9	9	4
(3)	왼쪽 손목	 수평 오른쪽으로 움직인다.	음색	음색- 멜로디의 Preset 음색을 바꾼다.	이전 프레임에 대한 손목의 방향	9	9	4
(4)	오른손	 수직 위아래로 움직인다.	멜로디	위로 움직이면 현재 음보다 높은, 아래로 움직이면 낮은 음의 멜로디를 선택한다.	-	-	-	-
(5)	오른손	 앞으로 뻗는다.	멜로디	선택된 멜로디 음을 재생한다.	-	-	-	-

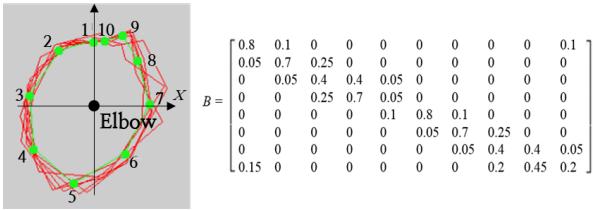


그림 3. 동작 (1)의 심볼과 확률분포 행렬

또한, 심볼은 그림 3 과 같이 6 번의 왼팔 원운동을 Record 한 뒤 그 위치의 평균을 구해 10 가지로 모델링 하였다. 그리고 이들 중 실제 왼쪽 손목의 팔꿈치에 대한 위치와 가장 가까운 관측 심볼을 현재의 관측으로 한다. 초기 심볼과 각 State $i(i=1, 2, \dots, 8)$ 에서 심볼 $j(j=1, 2, \dots, 10)$ 가 관측될 확률분포 행렬 $B=b(i,j)$ 는 그림 3 과 같이 초기화하였으며, 해당 동작의 관측열의 길이는 10 으로 설정하였다.

이와 같은 방법으로, 각각 동작의 HMM λ 에 대해, Forward Algorithm 을 이용하여 해당 동작이 관측될 확률을 구할 수 있고, 이 확률이 Threshold 값 이상이면 해당 동작을 수행했다고 판단하였다. Threshold 값은 동작에 따라 $10^{-5} \sim 10^{-6}$ 으로 정하였다. 또한, Baum-Welch Re-estimation Algorithm 을 이용해 사용자의 데이터를 기반으로 λ 를 학습시켰다.

2.2 화성 진행 및 멜로디 선택을 위한 알고리즘

본 시스템은 음악적 지식이 없는 사용자가 간단한 제스처를 이용하여 작곡할 수 있도록 하기 위하여, 대중 음악의 화성 진행을 분석하여 활용하였다. 서로 다른 장르의 대중 음악 7 개 곡을 마르코프 연쇄를 이용해서 분석하여 대중 음악에서 쓰인 화성 진행을 확률 모델로 얻었으며, 이를 기반으로 각 화음 바로 다음에 진행될 화음들의 확률을 구하고, 그래프 모델을 생성함으로써 자연스러운 화성 진행을 생성하였다.

본 연구에서는 하나의 코드를 나타내는 요소를 근음(Root) 및 코드의 종류로 추정하여 단순화 하였다.

- 근음 (12개) - 한 Octave 내의 음(C, C#, ..., B)
- 코드의 종류 (3개) - M(Major)/m(Minor)/S(7th)

일반적으로 대중음악에는 더 많고 복잡한 코드들도 사용하지만, 모두 이 두 가지 요소를 이용해 만들어질 수 있는 36 가지 코드들로 대체될 수 있다 [9]. 그러나 곡마다 조(Key)가 모두 달라서 단순히 코드를 이용한 마르코프 연쇄는 화성 진행으로서의 의미가 없다. 이를 해결하기 위해, 각 음악이 가지는 장조(Major



그림 4. C Major 와 D# Major Key 의 도수

Key)의 으뜸음(Key Note)부터 모든 음을 차례대로 1~12 의 도수를 매기고, 코드의 근음을 도수로 대체하였다. 예를 들어, C Major Key 와 D# Major Key 에서의 도수는 그림 4 와 같다. 본 연구에서는 이와 같이 근음의 음계 대신 도수를 사용한 화음을 도수 화음이라고 명명하고, 근음의 도수(1~12) 와 코드의 종류(M, m, S)를 붙여 써서 1M, 5m, 10S 와 같이 표기하였다. 예를 들어, 서로 다른 조(C Major Key, D# Major Key)로 쓰여진 동요 <Happy Birthday To You>의 화성 진행 표기를 비교하면 표 2 와 같다.

표 2. 동요 <Happy Birthday To You>의 화성 진행 표기

표기법	C Major Key	D# Major Key
코드 표기	C-G7-G7-C- C-F-G7-C	D#-B ^b 7- B ^b 7-D# -D#-G#-B ^b 7-D#
도수 화음 표기	1M-8S-8S-1M- 1M-6M-8S-1M	1M-8S-8S-1M- 1M-6M-8S-1M

따라서, 이 36 개의 도수 화음 각각을 마르코프 연쇄의 State 로 보고, 대중 음악의 화성 진행을 한 State 에서 다른 State 로 Transition 하는 것과 대응함으로써 모델링 하였다. 또한, 많은 음악에서는 4 개의 코드가 반복되는 경향이 있다. 이를 이용해 사용자가 화성 진행을 바꾸고자 할 때, 위에서 모델링한 마르코프 연쇄를 이용하여 새로운 4 개 화음의 화성 진행을 생성, 이를 반복하였다. 이 때 반복을 위해서는 마지막 화음과 처음 화음의 화성 진행 역시 자연스러워야 한다. 따라서, 그림 5 과 같이 생성한 화성 진행중 네 번째, 즉 마지막 화음을 첫

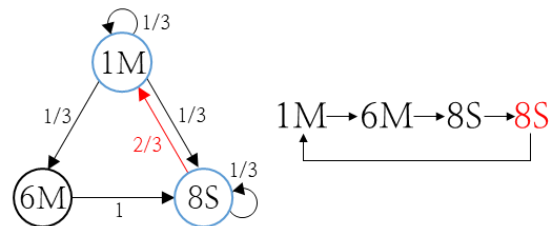


그림 5. 표 2 의 정보를 이용한 마르코프 연쇄와 생성된 화성 진행. 파란색 State 는 마지막에 올 수 있는 화음

번째 화음으로 Transition 할 수 있는 확률이 가장 큰 것으로 선정해주었다. 그림 5 는 표 2 의 정보를 이용해 마르코프 연쇄를 모델링하고 4 개의 코드를 생성한 것이다.

도수 화음으로 구성된 화성 진행을 소리로 출력하기 위해서 실제 코드로 다시 변환을 해야 하는데, 새로 만들어지는 음악의 조 (Key)를 정하면 도수 화음이 실제 코드에 대응이 된다. 즉 36 가지의 모든 실제 코드에 대한 소리만 있으면, 새로 작곡되는 곡의 조를 임의로 설정하여도 도수 화음에 해당하는 실제 코드의 소리를 출력해주기만 하면 되므로 작곡에 다양성과 자율성을 줄 수 있다.

멜로디는 오른손의 높낮이에 따라 결정되는데, 본 연구에서는 사람이 높낮이로 용이하게 구분, 표현할 수 있는 단계를 고려하여 최저 낮은 도(C3)부터 높은 도 (C5)의 음까지 총 25 가지 음을 사용하였다. 이 때 화음과 어울리는 멜로디의 선택을 위해 현재 재생되고 있는 화음과 불협화음을 내는 Avoid Note [10]를 제외시켰다. 예를 들어, 한 마디에서 C 화음이 연주될 때, C#4 같은 인접음, G#4 와 같은 불협화음을 만드는 음이 제외되는 식이다. 그리고 남은 음의 개수만큼 키넥트가 인식할 수 있는 높낮이 범위를 분할한 뒤, 오른손이 위치한 높이에 해당하는 음을 멜로디로 선택한다. 또, 오른쪽 어깨와 손의 깊이 차이로부터 사용자가 손을 뺐은 여부를 판별한 뒤 키넥트를 향해 오른손을 뺐으면 선택된 멜로디 음의 소리를 출력하였다.

3. 결론 및 향후 연구 계획

본 연구에서는 키넥트를 사용해, 전문적인 지식이 없는 비전문가도 두 팔과 손의 제스처를 이용해 음악을 손쉽게 작곡할 수 있는 NUI 기반의 시스템을 구현하였다. HMM 을 이용하여 사용자의 동작을 인식하고, 각 동작에 음악의 구성요소(멜로디, 리듬, 음색, 화성 진행)를 대응시켰다. 특히 화성 진행의 대응을 위하여 으뜸음을 기준으로 하는 도수 화음의 개념을 도입하였으며, 대중음악의 화성 진행의 특성을 활용하여 마르코프 연쇄를 모델링하였다.

차후에 음악 비전문가들을 대상으로 본 시스템을 사용한 뒤, 결과물에 대한 만족도와 흥미 등의 문항을 설문조사를 통하여 알아보고, 암맹 실험을 통하여 전문 작곡가의 저작과 비교할 계획이다. 또한, 현재는 대중음악의 분석을 통해 4 개의 화음으로 이루어진 화성 진행을 자동으로 만들어주기 때문에 보다 실험적이고 자율적인 작곡 결과물을 도출하기는 어려운데, 속도, 동작의 매끄러움, 크기 등 제스처의

여러 요소와 화성 진행과의 관계를 대응시킴으로써 이를 해결할 수 있을 것이다.

사사의 글

이 논문은 2017 년도 미래창조과학부의 재원으로, 한국연구재단의 기초연구사업 (NRF-2017R1A2B4008144), 정보통신기술진흥센터의 디지털콘텐츠원천기술개발 사업(2017-0-00179)의 지원으로 수행되었음.

참고 문헌

1. 윤성복, 정재훈, 안창욱. 유전 알고리즘을 이용한 원곡 기반의 음악 재생산 작곡 시스템. 한국정보과학회 학술발표논문집 (2014), 1791-1793.
2. 이운재, 김선민. 허밍을 이용한 고품질 음악 생성. 한국소음진동공학회 학술대회논문집 (2014), 146-149.
3. Simon, I., Morris, D., and Basu, S. MySong: Automatic Accompaniment Generation for Vocal Melodies. In Proc. SIGCHI Conference on Human Factors in Computing Systems, ACM (2008), 725-734.
4. Winkler, T. Making Motion Musical: Gesture Mapping Strategies for Interactive Computer Music. In Proc. 1995 International Computer Music Conference (1995), 261-264.
5. Kinect for Windows, <https://developer.microsoft.com/ko-kr/windows/Kinect/>
6. Hsu, M. H., Kumara, W. G. C. W., Shih, T. K., and Cheing, Z. Spider King: Virtual Musical Instruments Based on Microsoft Kinect. In Proc. 6th IEEE International Joint Conference on Awareness Science and Technology and Ubi-Media Computing (2013), 707-713.
7. Meyer, L. B. Explaining Music: Essays and Explorations. University of California Press (1973), 9-9.
8. Rabiner, L. & Juang, B. An Introduction to Hidden Markov Models. IEEE ASSP Mag., vol. 3 (1986), 4-16.
9. 백병동. 화성학. 제 12판. 서울: 수문당 (2008).
10. Levine, M. The Jazz Theory Book. O'Reilly Media, Inc. (2011).