Neural Network 기반 악기 보조 시스템

김대연, 오정록, 이수경, 강우철
인천대학교 임베디드시스템공학과
ssregibility@naver.com, king4719@naver.com, sing_815@naver.com,
wchkang@inu.ac.kr

A Neural Network Based Musical Instrument Support System

Dae Yeon Kim, Jeong Rok Oh, Soo Gyeong Lee, Woo Chul Kang
Embedded System Engeneering, Incheon National University

요 약

현재 초보적인 능력을 가진 악기 연주자가 접근할 수 있는 하드웨어, 소프트웨어를 사용해 악기 연주법을 연습할 수 있는 수단은 전무하다. 따라서 본 논문은 악기 연주자가 연습을 하기 위해 사용할 수 있는 음 인식과 악보 정보의 처리, LSTM을 통한 자동 악보 생성의 복합적 기능을 가진 악기 보조 시스템을 제안한다. 또한 본 시스템은 기존의 FFT와 같은 일반적인 Pitch Detection 알고리즘보다 더 우월한 음 인식 성능을 보유한 Autocorrelation 전처리를 거친 LeNet-5 Convolutional Neural Network 모델을 사용하여 음 인식 성능을 높이는 기법을 제안한다. 이 음 인식 모델은 실험 결과 기존의 음 인식 기법보다 최대 약 5.4%의 성능 증가를 이루어냈다.

1. 서론

기존의 악기 음 인식의 하드웨어와 소프트웨어적 분야는 튜너, 즉 사용자가 입력한 음을 인식하여 조율을 보조하는 측면에 한정되어있다.[1]

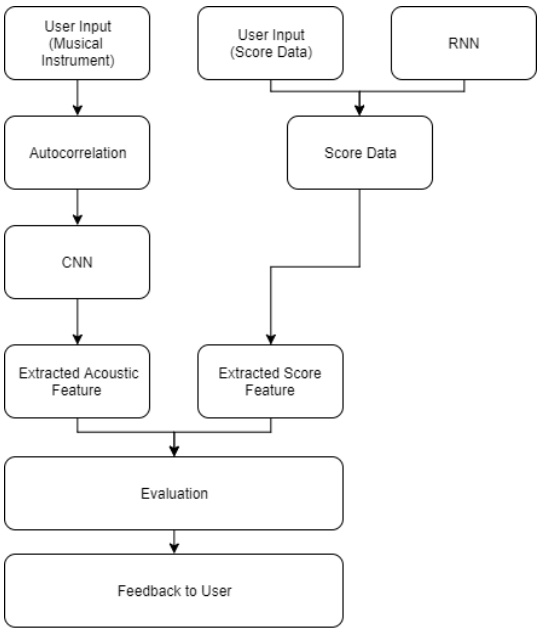
이는 즉 하드웨어와 소프트웨어를 사용한 사용자의 연주에 대한 정확한 평가를 하는 시스템의 부재를 의미하며, 실제로 현재 악기 연주법을 일반적인 사용자가 보유한 악보를 사용하여 배울 수 있는 방법은 실제 강사 또는 시청각 자료를 통한 연주법의 강의, 그리고 개인적으로 튜너와 악보를 통해 연주법을 스스로 익히는 방법이 유일한 접근 방식이다.

이러한 현 상황을 개선하기 위해 음의 인식과 임의의 악보 정보의 대조를 통한 유사한 보조 체계에 대한 유사한 여러 제안[2]이 존재하지만, 인식 성능이 떨어지는 FFT를 사용하는 경우가 많고, FFT를 사용하지 않아도 Neural NEtwork를 사용하지 않기 때문에 사용자 입력 음의 인식 성능과 인식성능을 개선할 여지가 있다.

따라서 이 한계점을 해결하기 위해 본 논문은 현재 기타 튜너의 음 인식에서 주로 사용되는 기법인 FFT보다 우수한 Autocorrelation[3]에 기반한 음 인식을 제안한다. 또한 Neural Network를 사용하여, 단순한 알고리즘보다 더 높은 정확도로 사용자 임의의 악보와 연주의 일치 정도를 평가 하는 시스템을 제안한다. 마지막으로 사용자의 연주 능력을 향상시키기 위해 Recurrent Neural Network를 통해 무작위적으로 생성된 악보에 대한 사용자의 연주 능력을 평가하는 체계적인 시스템을 제안한다.

본 논문에서 제안한 Autocorrelation과 Neural Network의 조합은 대조를 위해 작성된 FFT 기반의 Neural Network와 단순 알고리즘들보다 훨씬 향상된 소리신호를 인식하는 능력을 보여주었다.

2. 시스템 구조



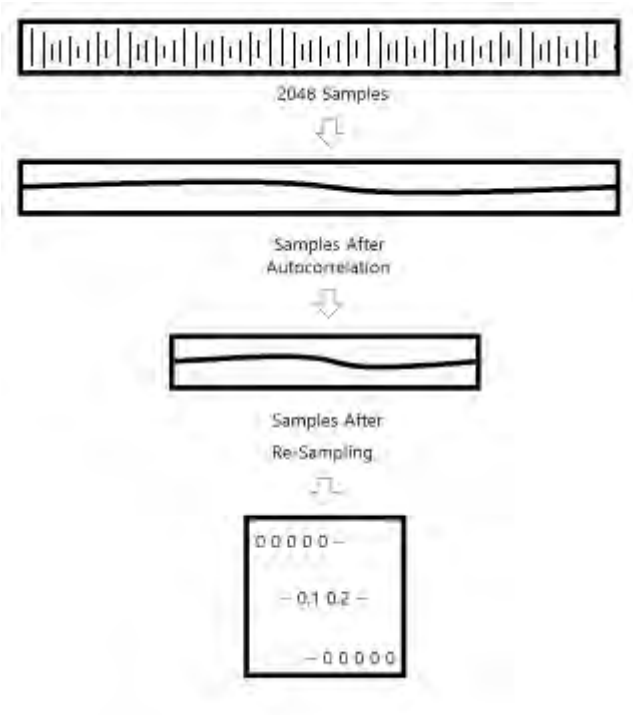
(그림 1) 시스템 구조

그림 1은 Neural Network에 기반한 악기 보조 시스템의 개요도이다. 입력된 악기 신호는 Autocorrelation을 통해 변환된 이후 이 결과가 Convolutional Neural Network에 입력되어 최종적으로 음의 높이로 정의된 Class의 값이 얻어진다.

한편, 유저가 선택한 임의의 악보 데이터 또는 Recurrent Neural Network를 통해 생성된 무작위적인 악보 데이터는 시간 단위로 데이터를 대조하여 현재의 연주 시간과 사용자의 입력, 그리고 악보의 정보가 일치하는가 여부를 판단하여 사용자에게 실시간적인 피드백을 제공한다.

3. 시스템 구현

3.1 User Input



(그림 2) 소리 입력의 전처리 과정

사용자의 입력은 악기 입력과 악보 입력으로 이루어지며, 두 개별적인 입력 자료를 대조가 가능한 형태로 변환하고, 추가적으로 유저의 입력에 상응하는 시간 역시 저장한다.

악기 입력의 경우, 마이크를 사용하는 경우에는 주변 환경의 소리가 잡음으로 작용하여 인식 능력을 떨어트릴 수 있기에 이 잡음을 최소한으로 줄이기 위해 USB 포트에 연결이 가능하도록 A/D 컨버터가 사용된 모노 6.5mm 케이블을 사용한다.

44100Hz의 Sampling Rate로 입력된 소리 신호의 Stream은 실시간성을 유지하기 위해 적은 시간동안의 입력만을 고려해야 하며, 따라서 2048 Sample 입력을 단위로 하여 Autocorrelation을 적용한다.

Autocorrelation이 적용된 이후에는 총 1024개의 Sample이 남도록 균등하게 다시 Sampling을 거친 이후, 최종적으로 32x32 Matrix로 변환한다.

또한 매 유저 입력에 대한 시간을 측정하기 위해, 위의 악기 입력과 전처리 과정은 인간이 한 음을 연주할 수 있는 시간보다 훨씬 짧은 0.05초에 한번씩 진행하여, 1회의 처리마다 0.05초가 소모된 것으로 처리하여 시간의 진행을 기록한다.

악보 입력의 경우, 기존에 존재하는 악보 데이터와의 호환성, 그리고 사용자의 유동적이고 임의적인 악보의 선택을 보장하기 위해 흔히 사용되는 Guitar Pro 소프트웨어의 악보 파일 포맷을 사용하여 정보를 읽고, 악보에 대한 정보를 취득한다. 악보에 저장된 정보는 음의 상대적 높이로 표현된 1byte의 정수 데이터와 각 해당하는 음에 대한 절대적인 시간을 float형 데이터의 리스트 형태로 저장한다.

3.2 Recurrent Neural Network

프렛 - 현 위치	길이	추가 정보

(그림 3) 개별 음의 텍스트 포맷

사용자가 악보를 입력하지 않고 무작위적인 악보를 요구할 때 악보를 생성하기 위해 Recurrent Neural Network는 LSTM[4]을 사용한다. Guitrar Pro 포맷의 악보 파일의 읽고 이 악보 파일의 음악적 정보를 개별 음 단위로 분리하여 음의 프렛과 현 위치, 길이, 그리고 길이의 가변성과 같은 정보를 담은 가변적 길이의 텍스트 양식(그림 3)으로 포맷화한 뒤, 이와 같은 포맷화 텍스트의 빈도를 사용해 개별적인 구조를 가진 각 음의 출현 빈도를 고려하여 전반적인 음이 구성되는 평균적 형태와 특이한 형태가 구분되어 반영하도록 학습한다.

LSTM의 학습을 통해 생성된 결과는 포맷에 담긴 정보를 분석하여 사용자의 Guitar Pro 포맷 파일의 입력과 동일한 리스트형의 자료로 저장한다.

3.3 Convolutional Neural Network

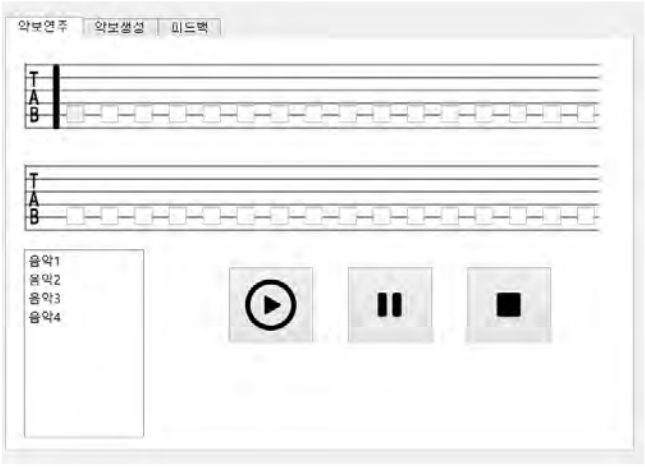
소리의 특징을 학습하기 위한 Convolutional Neural Network는 LeNet-5[5] 모델을 사용한다. Convolutional Neural Network는 2012년 AlexNet의 성공적 결과 이후부터 Visual Recognition 분야의 정석적인 방법으로 인지[6]되고 있으며, Classification의 대상인 특정한 Feature를 지닌 이미지가 입력으로 주어진다는 특성은 Autocorrelation을 거친 소리 정보가 주파수적 특징을 드러낸다는 점에서 큰 유사성과 호환성을 가지고 있기 때문에 Convolutional Neural Network를 사용한다.

3.4 User Input Evaluation

사용자의 입력 평가는 Convolutional Neural Network를 거쳐 특정한 음으로 Classification을 거친 정보와 현재 연

주의 진행 시간, 그리고 악보의 각 개별적인 음에 대한 음의 높이와 시간의 정보가 List 형태로 저장된 Score Data 세 요소의 대조를 사용하여 정확하게 연주되었는지, 정확하게 연주되지 않았는지를 이진적으로 평가한다.

3.5 User Feedback



(그림 4) 연주 평가 GUI

사용자에게 현재 연주에 대한 실시간적인 평가를 제공하기 위해 GUI를 사용하여 사용자가 입력한 임의의 악보의 정보를 타브 악보 형태로 표시하고, 현재 연주가 진행되고 있는 악보에서의 위치, 그리고 각 음에 대해서 사용자가 정확히 연주하였는가에 대한 정보를 시각적으로 제공한다.

4. 실험결과

	Accuracy	Excessive delay
FFT	~92%	No
Autocorrelation	~96%	No
FFT with Linear Regression	73.1%	No
FFT with LeNet-5	93.7%	No
Autocorrelation with LeNet-5	97.4%	No
FFT와 Autocorrelation with LeNet-5의 비교	약 5.4% 증가	No

<표 1> 실험 결과, 본 논문에서 제안하는 모델은 Autocorrelation with LeNet-5이다.

실험은 악기 입력에 대해 다른 여러 소리 신호의 처리 방식을 사용한 결과와 본 논문에서 사용한 처리 방식인 Autocorrelation 전처리를 거치는 LeNet-5 모델의 성능 비교, 그리고 테스트시 사용한 악보의 일부와 난수를 사용해 생성된 무작위적 악보, 그리고 LSTM을 통해 생성된 악보를 대조하였다.

본 논문의 LeNet-5 모델은 4개의 현의 15개의 프렛에 대해 3분동안 녹음한 음을 학습 자료로 사용하여 총 31개의 Class에 대해 학습을 진행하였으며, 다른 모델의 정확도는 이 학습 자료의 일부를 무작위 추출하여 평가하였다.

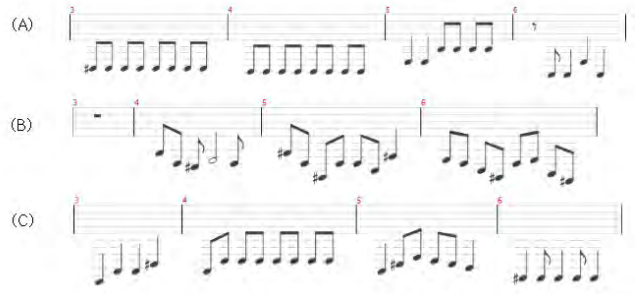
LSTM 학습은 균일성을 유지하기 위해 동일한 음악 장르와 동일한 음악가가 작곡한 5개의 악보를 사용하였다.

<표 1>은 소리 신호의 처리 방식에 대한 결과이다. 학습에 사용된 자료는 음이 약해지는 구간 역시 포함하기 때문에, FFT와 Autocorrelation의 경우에는 여러 차례의 테스트를 거친 후 대략적인 평균치를 사용하였다.

모든 처리 방식은 악기신호 입력 Stream의 2048 Sample (약 0.05초)이 받아지는 시간과 유사한 시간 내로 과도한 지연시간 없이 성공적으로 음의 처리와 인식을 성공했으며, FFT보다 Autocorrelation을 사용한 경우가 더 높은 인식 성능을 보여준다.

또한 LeNet-5 Convolutional Neural Network를 사용한 경우 사용하지 않고 일반적인 Pitch Detection 알고리즘을 사용하거나 Linear Regression을 사용한 경우에 대해서 정확도의 증가를 보여주었으며, 본 논문의 Autocorrelation을 사용해 전처리한 데이터를 학습한 모델은 전처리를 FFT로 교체하여 사용한 데이터에 비해서 더 높은 인식 성능을 보여준다.

최종적으로, 기존의 FFT 알고리즘에 비해 본 논문의 Autocorrelation을 사용한 LeNet-5 모델은 약 5.4%의 인식률 증가를 보여준다.



(그림 5) 생성된 악보의 대조

(그림 5)는 악보 생성의 결과를 보여준다. (A)의 경우 악보 생성을 위한 LSTM 모델에 사용된 학습 자료의 일부이다.

(B)의 경우 무작위적인 난수를 통해 생성된 악보의 일부이다. 이 경우 리프 구조에 대한 고려가 반영되지 않고 무작위적으로 생성된 결과를 볼 수 있다.

(C)의 경우 학습된 LSTM을 통해 생성된 악보의 일부이다. (B)와 달리 (A)의 경우와 같이 리프 구조와 개별 음의 음정과 길이에 대한 규칙성이 반영되어 생성된 결과를 볼 수 있다.

5. 결론 및 향후 연구

현재 초보적인 실력의 악기 연주자가 접근할 수 있도록 간편한 동시에 사용자가 원하는 임의의 악보를 사용해 연주를 평가할 수 있는 시스템은 존재하지 않는 실정이다. 본 논문에서는 이에 따라 사용자의 악기 입력 신호와 사용자가 지정한 임의의 악보에 대응하여 연주를 평가하고 피드백을 줄 수 있는 Neural Network 기반 악기 연주 보조 시스템을 제안한다.

Neural Network 기반 악기 연주 보조 시스템은 단순한 알고리즘을 통해 음을 인식하는 것보다 더 우수한 성능을 가진다는 것을 실험을 통해 확인한 Autocorrelation 전처리를 거치는 Convolutional Neural Network를 통한 입력 신호와 악보 정보의 대조, 사용자가 지정한 악보 외에도 무작위적으로 생성된 악보를 통해 연주에 대한 피드백을 얻을 수 있도록 하는 Recurrent Neural Network를 통해 악보를 생성하는 기능을 통합한 시스템이다.

해당 논문 연구 결과는 음 인식의 경우 Spectral Flux[7]와 같은 일반적으로 사용되지 않는 여러 다양한 종류의 Pitch Detection 알고리즘의 전처리 방식에 대해 이와 같은 Convolutional Neural Network를 적용할 시의 성능 개선 여부와 AlexNet[6]과 같은 더 깊고 복잡한 네트워크 모델이 적용된 경우와 현재 모델의 성능 비교의 측면에서 추후 연구가 가능하다.

또한 현재 사용된 Autocorrelation 알고리즘의 고유한 한계점과 Neural Network를 사용하여 발생한 기술적 문제로 인한 복합적인 한계점인 화음을 분류할 수 없는 문제의 해결에 대해서는 개선의 여지가 있다.

마지막으로 Recurrent Neural Network를 통해 생성하는 무작위적 악보는 짧은 시간 단위를 중점에 두고 작성되었기 때문에, 음악적 구조 역시 학습하여 좀 더 길고 복잡한 무작위 생성 악보를 만드는 방법에 대해 추가적인 개선과 연구가 가능하다.

6. Acknowledgement

본 논문은 송도산업단지캠퍼스조성사업단의 지원을 받아 작성되었음.

본 논문은 2017년 한이음 ICT멘토링 프로젝트의 결과물입니다.

참고문헌

[1] Devansh Zurale, Harshal Tankaria, Meghna Mandalia, Raj Sheth. Automatic Guitar Tuner.

[2] Yoonchang Han, Sejun Kwon, Kibeom Lee, Kyogu Lee. A Musical Performance EvaluationSystemfor Beginner Musician based on Real-time Score Following.

[3] L. Ihsan Yazici, Mursel Onder, Aydin Akan. Recognition of Monophonic Musical Notes Using Short-time Autocorrelation Estimate.

[4] Y. LeCun and Y. Bengio. Convolutional networks for images, speech, and time-series. In M. A. Arbib, editor, The Handbook of Brain Theory and Neural Networks. MIT Press, 1995.

[5] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.

[6] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks.

[7] Dixon, S., Onset detection revisited. Proceedings of the 9th International Conference on Digital Audio Effects. Vol. 120. 2006.