가상현실 음향 재구성을 위한 심층신경망 기반 스마트폰의 주파수응답 정규화 방법

최재규 윤덕규 최승호
서울과학기술대학교
shchoi@snut.ac.kr

A Method of Frequency Response Normalization of Smart Phones Based on Deep Neural Networks for Virtual Reality Sound Reconstruction

Jaegyu Choi Deokgyu Yun Seung Ho Choi
Seoul National University of Science and Technology

요약

본 논문은 사용자제작콘텐츠 (User Created Contents, UCC)를 이용한 가상현실 (virtual reality, VR) 음향 재구성을 위한 스마트폰의 주파수응답 정규화에 관한 연구이다. 서로 다른 스마트폰들로 취득한 음향들을 연결할 때, 부자연스러운 음향이 발생하며, 이것은 주로 스마트폰별로 다른 주파수 응답에 기인한다. 이러한 문제를 해결하기 위하여 스마트폰의 주파수응답을 정규화 하는 과정이 필요하며, 본 연구에서는 심층신경망 (deep neural network)을 이용하는 방법을 제안한다. 심층신경망의 입력은 처리하고자 하는 스마트폰 음향의 스펙트럼이고 출력은 이것과 기준 스마트폰 음향의 스펙트럼과의 비율이다. 실험결과, 서로 다른 스마트폰으로 취득한 음향신호가 연결되었을 때, 객관적 및 주관적 평가를 통해 음향의 자연성이 개선됨을 확인하였다.

1. 서론

사용자 제작콘텐츠 (User Created Contents, UCC)를 이용한 가상현실(virtual reality, VR) 미디어를 제공하고자 할 때, 기존의 VR 미디어는 파노라마 카메라 또는 360도 VR카메라 같은 특별한 장비로 취득되어 제작되었으며, 사용되는 음향 장비 또한 특수한 마이크였다. 이러한 방법에는 전문가들이 직접 촬영 및 녹취를 한 미디어만이 제공될 수 있다는 한계가 있으며 따라서 그러한 장비를 사용하지 않은 대다수의 UCC에 기존 VR 기술을 적용하기에는 적합하지 않다. UCC를 사용해 VR에 적합한 미디어를 제공하기 위해서는, 다수의 일반 관람객들이 각자의 스마트폰으로 촬영한 영상 및 오디오 신호만을 가지고 실감미디어를 만들어야 한다. 이를 위해선 서로 다른 위치에서 받은 다양한 영상의 정합 처리[1]가 필요할 뿐만 아니라, 다양한 시점과 시간에서 취득한 음향에 대한 동기화 및 스마트폰 기종 차이에 의한 주파수 응답의 정규화 등 음향처리도 필수적이다.

이러한 문제를 해결하기 위하여 본 논문에서는 심층신경망 (deep neural network)을 이용하여 각기 다른 스마트폰으로 취득한 동일한 음향신호를 정규화 하는 방법을 제안한다. 서로 다른 주파수 응답을 가진 음향 신호의 스펙트럼을 입력하여 기준 기종의 것과 유사한 스펙트럼이 되도록 하는 비율을 출력하도록 심층신경망을 훈련시킨다. 이러한 과정을 통해 다른 기종에서 취득한 음향을 부드럽게 연결하여 자연성을 개선할 수 있다.

2. 스마트폰 기종에 따른 주파수 응답 비교

먼저 각기 다른 스마트폰 기종의 주파수 응답을 비교하기 위하여 시간에 따라 주파수가 증가하는 정현파 신호 (sine sweep signal)를 이용하여 주파수 응답을 분석해 보았다. 그림 1은 4가지 기종에 해당하는 주파수 응답이다. 보이는 바와 같이 차단 주파수와 응답 특성이 서로 다른 것을 알 수 있다. 본 연구에서는 다른 스마트폰들의 차단 주파수를 고려하여 iPhone 7을 기준 스마트폰으로 정하였다. 따라서 심층신경망의 출력은 입력신호에 대한 iPhone 7에 해당하는 단구간 스펙트럼의 비율로 설정한다.

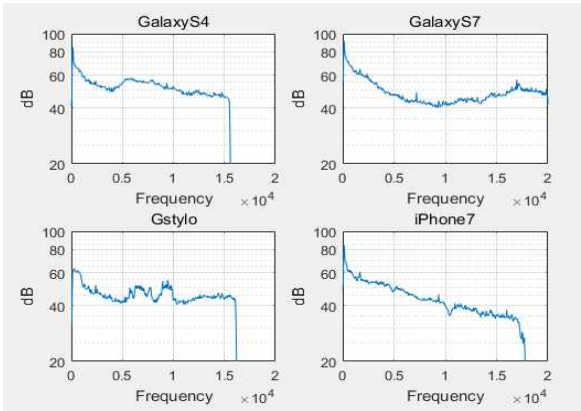

그림 1. 스마트폰 기종별 주파수 응답

3. 심층신경망 기반 주파수응답 정규화

여러 스마트폰 기종의 스펙트럼이 입력되면, 기준 스마트폰 기종의 스펙트럼이 되는 비율을 출력하도록 심층신경망을 훈련시킨다. 활

성함수 (activation function)로는 노드의 입력 x에 노드 출력이 max(0, x)인 ReLU (Rectified Linear Units) [2]를 사용하였으며, 활성함수 이후에 drop out [3]도 사용하였다. 또한 학습시 최적화 알고리즘은 ADAM (ADAptive Moment estimation) [4]을 사용하였다. 전체 구조는 그림 2와 같다.

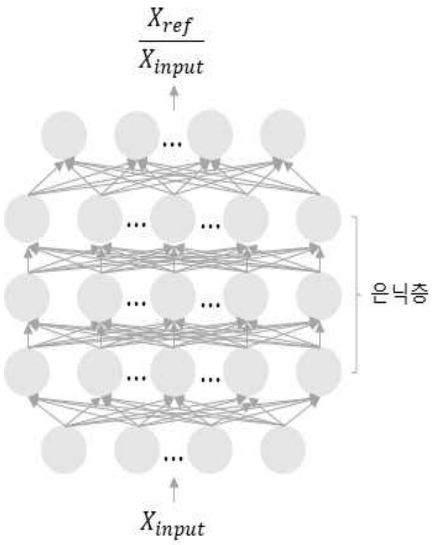

그림 2. 심층신경망의 구조

그림 2에서 X_{input}은 여러 기종의 스펙트럼이며, X_{ref}는 기준 (reference) 스마트폰의 스펙트럼이다. 이렇게 훈련된 심층신경망을 통해 각기 다른 스마트폰의 스펙트럼 비율의 추정치를 구하고, 그 추정치를 입력 스펙트럼에 곱하여 최종 처리된 출력 음향을 구함으로써 스마트폰의 주파수응답 정규화가 이루어진다.

4. 실험 및 결과

먼저 다수의 스마트폰과 iPhone 7으로 약 30분간 여러 가지 음향을 취득하였으며, 샘플링 주파수는 44,100 Hz이다. 이렇게 취득한 데이터로 iPhone 7로 취득한 데이터와 동기화 처리를 한 후에 단구간 스펙트럼을 구하였다. 프레임 크기는 20 ms이고 1024 포인트 FFT를 사용하였으며, 50 % 중첩하였고 Hanning 윈도우를 사용하였다. 따라서 심층신경망을 구성할 때 입력 노드와 출력 노드의 수는 주파수 해상도 (frequency bin)에 해당하는 513개가 된다. 실험에 사용한 심층신경망의 은닉층은 개수가 3층이고 노드의 수는 각각 1024이다.

그림 3에 iPhone 7으로 취득한 신호의 스펙트로그램과 여러 기종의 스마트폰으로 취득한 신호를 1초 간격으로 연결한 것, 그리고 심층신경망으로 연결한 음향을 정규화 처리한 신호의 스펙트로그램을 나타냈다. 처리 전의 스펙트로그램에서 관측되는 기종별 차단주파수 차이 때문에 계단식으로 보이는 구간들과 경계구간에 있는 노이즈 등이 처리 후에는 많이 보완되었으며, 전체 스펙트로그램이 iPhone 7의 스펙트로그램과 비슷해지고 비공식적 청취 테스트 (informal listening test)를 통해 소리가 자연스러워졌음을 확인하였다. 객관적인 수치로 결과를 비교하기 위하여 (Log Spectral Distance, LSD)를 계산하였다. iPhone 7과 처리 전의 거리 8.19 dB에서 처리후 6.33 dB로 감소하였다.

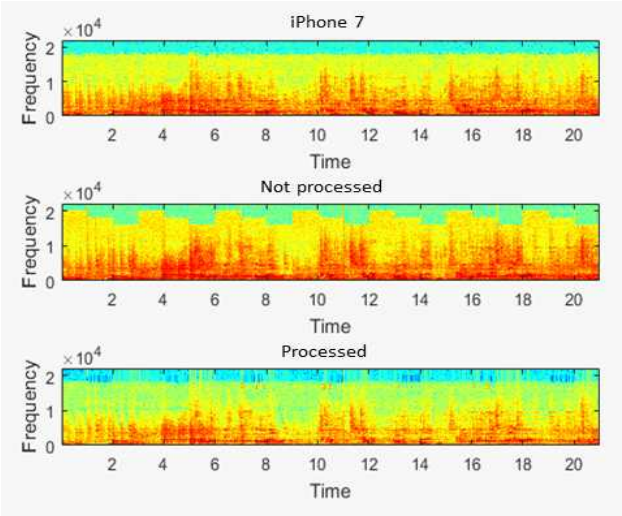

그림 3. 처리 전과 후의 스펙트로그램 변화

5. 결론 및 향후 연구방향

본 논문에서는 UCC를 이용한 VR 음향구현에 있어서 스마트폰 기종에 따른 스펙트럼의 차이를 보상하기 위해 심층신경망을 이용해 주파수 응답을 정규화 하는 방법을 제안하였다. 실험 결과, 서로 다른 기종으로 취득한 음향신호들을 연결할 경우에, 스펙트로그램과 LSD 거리 감소를 통해 효과가 있었음을 보였다. 향후 고성능 마이크로폰의 주파수응답을 기준으로 하여, 정규화뿐만 아니라 음질개선을 위한 연구를 진행하고자 한다.

감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. [2016-0-00144, 시청자 이동형 자유시점 360VR 실감미디어 제공을 위한 시스템 설계 및 기반기술 연구]

참고문헌

[1] 이동석, 유지환, 유지상, “히스토그램 매칭을 이용한 다시점 비디오의 휘도와 색차 성분 보상 기법,” *한국방송미디어공학회 학술발표대회 논문집,* 191-194, 2009.

[2] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in *Proc. 27th Int. Conf. Machine Learning*, pp. 807-814, 2010.

[3] Vu Pham, Théodore Bluche, Christopher Kermorvant, and Jérôme Louradour, “Dropout improves recurrent neural networks for handwriting recognition,” *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference,* pp. 285 - 290, IEEE, 2014.

[4] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.