조건부 오토 인코더를 이용한 오디오 고대역 부호화 기술

*조효진 **백승권 *장원 *신성현 *박호종

*광운대학교 **한국전자통신연구원

*got7ze@kw.ac.kr

High-Band Coding of Audio Signal Based on Conditional Auto Encoder

*Cho, Hyo-Jin **Beak, Seung-Kwon *Jang, Won *Shin, Seong-Hyeon *Park, Hochong

*Kwangwoon University, **Electronics and Telecommunications Research Institute

요약

본 논문에서는 조건부 오토 인코더를 사용하여 오디오 고대역 신호를 부호화 하는 기술을 제안한다. 오토 인코더의 데이터 압축 특성을 이용하여 부호화를 위한 데이터의 양을 크게 줄인다. 제안하는 알고리즘은 기존의 오토 인코더와 달리 과거의 정보가 포함된 2차원 조건을 함께 입력하여 오토 인코더가 코딩 프레임의 고대역을 복원하는 것을 돕도록 한다. 2차원 조건과 입력을 압축하여 연결한 후 디코딩하여 코딩 프레임의 고주파 대역을 만든다. 제안하는 방법을 사용하면 저대역 MDCT 계수와 고대역 MDCT 계수를 오토 인코더로 압축한 결과만으로 원본과 유사한 음질을 청취할 수 있다.

1. 서론

한정적인 채널 용량과 메모리 크기에 대비하여 디지털 오디오는 저장 및 전송 시 데이터양을 줄이기 위해 부호화된다. 효과적인 통신과 메모리 관리를 위해서는 부호화 기술의 압축률이 높을수록 유리하지만, 음질의 손상이 심하다면 압축률이 높아도 부호화의 의미가 없다. 높은 음질과 압축률 두 가지를 모두 만족하는 부호화 기술은 꾸준히 연구되어 왔으며, 최근까지 가장 높은 효율을 보인 부호화 기술은 파라메트릭 (parametric) 기술로 우수한 음질을 보장하면서도 높은 압축률을 갖는다. 파라메트릭 기술이란 전송할 정보를 소수의 파라미터로 표현하고, 양자화 한 파라미터를 전송하고, 복호기에서 전송된 파라미터로부터 정보를 복원하는 기술이다. 파라메트릭 기술의 예로, 오디오 고대역의 밴드 에너지만을 전송하여 저대역 데이터와 고대역의 밴드 에너지만으로 고대역의 데이터를 생성하는 방법 등이 있다[1].

파라메트릭 기술은 일정 압축률에서 높은 음질을 자랑하지만 압축률은 한계가 있다. 파라메트릭 기술의 압축률은 전송하는 파라미터의 수에 의해 정해진다. 높은 압축률은 복원하는데 필요한 정보를 덜 전송한다는 것이며 음질의 손상을 의미한다. 그러므로 특정 압축률 이상으로 압축률이 높아지면 원본에 비해 음질이 저하된다. 이러한 기존 기술의 한계를 극복하기 위해 오토 인코더 (auto encoder)를 이용하여 오디오를 부호화하는 알고리즘을 제안한다[2].

오토 인코더는 입력과 출력이 같은 비지도 (unsupervised) 학습 신경망으로 특정 입력 A가 입력되었을 때의 출력 $\hat{A}$이 입력 A와 같아지는 방향으로 학습한다. 입력의 크기보다 작은 은닉 뉴런의 수로 입력을 압축할 수 있으며 충분한 학습이 이루어진다면 큰 압축률에도 높은 음질을 유지할 수 있다.

본 논문에서는 오토 인코더의 현재 입력뿐만 아니라 과거의 정보가 포함된 2차원 조건 (condition)을 입력하여 오토 인코더를 훈련하는 알고리즘을 제안한다. 모든 동작은 MDCT (modified discrete cosine transform) 차원에서 진행되며 복원의 목표가 되는 프레임을 코딩 프레임으로 정의한다. 제안하는 알고리즘을 사용하면 저대역의 MDCT 계수와 코딩 프레임의 고대역 MDCT 계수를 오토 인코더를 이용하여 높은 압축률로 압축한 결과만을 가지고 원본에 가까운 오디오 고대역을 복원할 수 있다.

2. 제안하는 방법

본 논문에서는 기존 오토 인코더의 입력뿐만 아니라 신경망의 학습을 도울 2차원 조건을 활용하는 조건부 오토 인코더 (conditional auto encoder)를 사용한다. 입력은 코딩 프레임의 고대역을 나타내는 MDCT 계수이다. 입력을 압축하기 위해서 DNN (deep neural network)을 사용하였으며 고대역을 표현하는 MDCT 계수를 압축하여 파라미터 벡터 I를 만든다[3].

조건은 코딩 프레임 이전 15개 프레임의 저대역과 각 프레임의 복원된 고대역 MDCT 계수 그리고 코딩 프레임의 저대역 MDCT 계수로 구성된다. 이전 프레임들의 전체 주파수 대역을 통해 고대역과 저대역의 관계를 추정할 수 있고, 추가 조건으로 입력하는 코딩 프레임의 저대역으로부터 고대역을 추정하는 데 도움이 되고, 이전 프레임들의 고대역을 통해 현재 고대역의 시간적 변화를 신경망에 학습시킬 수 있다. 이를 CNN (convolutional neural network)을 이용하여 파라미터 벡터 C로 압축한 후 I와 연결하여 디코더에 입력한다[4].

디코더는 입력을 압축하는 신경망과 동일하게 DNN으로 구현하였다. 디코더에서 출력된 값들을 원본 저대역 MDCT 값들과 연결한 후 IMDCT (inverse MDCT)를 거쳐 신호를 복원한다. 그림 1은 본 논문에서 사용한 조건부 오토 인코더 구조를 보여준다. 이와 같이 인코더 2개와 한 개의 디코더로 이루어진 Y형 구조로 신경망을 설계하였다.

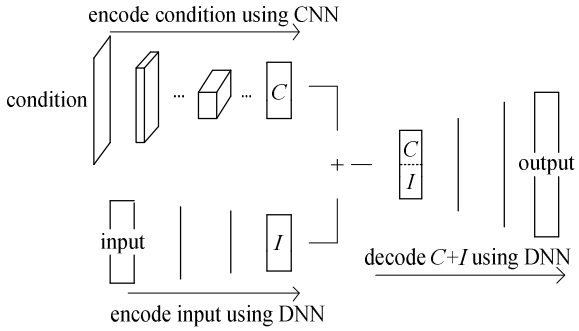

그림 1. 조건부 오토 인코더 구조
Fig 1. Structure of conditional auto encoder

3. 성능 평가

객관적 성능 평가를 위해 스펙트로그램 비교를 진행하고, 주관적 평가를 위해 청취 평가를 진행하였다. 각 성능 평가에는 저대역 정보만을 가지는 신호 (low-passed), 제안하는 알고리즘을 사용하여 복원한 신호 (proposed), 원본 신호 (original)를 실험 데이터 세트로 사용하였다. 제안 알고리즘은 5개의 파라미터를 전송한다.

그림 2는 실험 데이터 세트의 일부 스펙트로그램을 나타낸 것이다. 그림 2-(a)와 같이 고대역 정보가 없는 신호로부터 제안하는 알고리즘을 사용하여 고대역을 복원한 결과 그림 2-(b)와 같은 결과를 얻었다. 그 결과가 원본의 스펙트로그램을 나타내는 2-(c)와 흡사함을 확인할 수 있다.

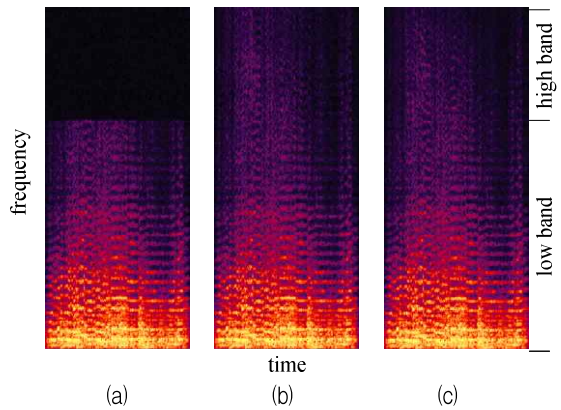

그림 2. 실험 데이터 세트의 스펙트로그램 (a) 저대역 정보만 가지는 신호, (b) 제안하는 알고리즘을 사용하여 복원한 신호, (c) 원본 신호
Fig 2. Spectrogram of test data set (a) low-passed, (b) proposed, (c) original

주관적 평가를 위한 청취 평가는 원음 신호와 테스트 신호 사이의 음질 차이를 측정하는 상대 음질 평가로 진행하였다. 원음 신호와 저대역 정보만을 가지는 신호, 그리고 원음 신호와 제안하는 알고리즘을 사용하여 고대역을 복원한 신호를 비교하였고, 평가 방법으로는 청취 평가에 자주 쓰이는 AB7을 사용하였으며 6명을 대상으로 진행하였다. 점수가 0에 가까울수록 테스트 신호와 원본 신호가 비슷한 음질임을 뜻한다. 그림 3이 청취 평가 결과이며, 각 항목의 가운데 선은 평균값을 나타내고 위, 아래 선은 95% 신뢰구간을 나타낸다.

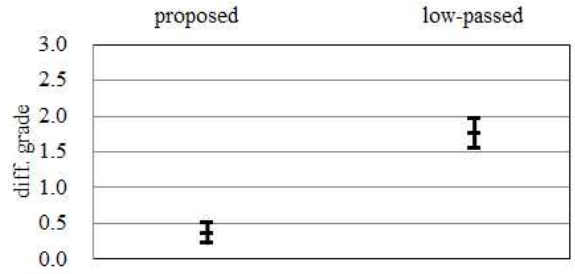

그림 3. 제안한 알고리즘으로 복원한 신호와 저대역 통과 신호의 AB7 청취 평가 결과
Fig 3. Results of AB7 test for the decoded signal by the proposed algorithm and the low-passed signal

그림 3에서 보듯이, 제안한 방법으로 복원된 신호의 평균 점수가 0과 가깝고 두 결과의 신뢰구간이 겹치지 않는 것을 확인할 수 있다. 이를 통해 제안하는 알고리즘이 고대역을 효과적으로 복원하여 음질을 향상시키는 것을 확인할 수 있다

4. 결론

본 논문에서는 조건부 오토 인코더를 이용한 오디오 고대역 부호화 알고리즘을 제안하였다. 과거의 정보와 코딩 프레임의 저대역 MDCT 계수로 구성된 조건을 추가로 입력하여 신경망을 학습시켰다. 제안한 알고리즘을 사용하면 오토 인코더를 이용하여 고대역을 표현하는 MDCT 계수를 크게 압축할 수 있다. 고대역을 표현하는 MDCT 계수의 개수를 크게 줄였지만 음질이 원본과 유사함을 성능 평가 결과를 통해 확인할 수 있었다. 심화연구를 통하여 저대역의 기준을 낮춰 전송하는 데이터의 큰 비중을 차지하는 저대역 데이터양을 더 줄일 수 있을 것으로 기대된다.

감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

참고문헌

[1] C. R. Helmrich, A. Niedermeier, S. Disch and F. Ghido, “Spectral envelope reconstruction via IGF for audio transform coding,” *Proc. IEEE ICASSP-15*, Brisbane, April 2015.

[2] G. E. Hinton and R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504 - 507, 2006.

[3] Bengio Yoshua, LeCun Yann, Hinton Geoffrey, "Deep Learning," *Nature*, vol. 521, pp. 436 - 444, May 2015.

[4] K. J. Piczak, “Environmental sound classification with convolutional neural networks,” in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1-6, Sep 2015.