잔차 신호 복제 기반 오디오 대역 확장 방법

임우택, 백승권, 이태진
한국전자통신연구원 미디어부호화연구실
wtlim@etri.re.kr

Research on audio bandwidth extension using residual signal replication

Wootaek Lim, Seungkwon Beack, Taejin Lee

Media Coding Research Section
Electronics and Telecommunications Research Institute (ETRI)

요 약

오디오 대역 확장 기술은 저 해상도의 오디오 신호를 고 해상도의 오디오 신호로 복원 또는 생성해 내는 기술이다. 이와 관련하여 오디오 코덱에서는 고 대역 오디오 신호의 저 비트 부호화를 위해 사람이 청각이 둔감하게 인지하는 고 대역의 오디오 신호에 대해 실제 신호에 대한 양자화를 수행하지 않고, 코딩 되어 전송된 저 대역 신호와 고 대역의 파라미터를 이용하여 신호를 합성하는 스펙트럼 대역 복제 기술이 널리 사용된다. 본 연구에서는 선형 예측 기반의 주파수 대역 복제 방법을 통해 추가 정보를 활용한 오디오 대역 확장을 수행하고 신경망 기반의 오디오 신호 개선을 통해 복제된 신호의 개선 가능성을 검토하였다. 실험 평가는 MPEG 에서 코덱 평가용으로 사용되는 테스트 시퀀스를 사용하였으며, 실험 결과 제안하는 방법을 적용하여 기존 오디오 대역 확장 기술 대비 성능이 향상됨을 확인하였다.

1. 서론

최근 고품질의 미디어 서비스를 위하여 다양한 고해상도 디스플레이나 고품질 오디오를 재생하는 디바이스들이 널리 보급되고 있으며, 이와 함께 고 해상도의 콘텐츠를 소비하고자 하는 사용자들의 요구사항 또한 증가하고 있다. 이에 따라 비디오와 오디오가 포함된 미디어 분야에서도 낮은 해상도의 콘텐츠를 높은 해상도의 콘텐츠로 향상시키고자 하는 연구가 진행되고 있으며, 대표적인 연구 분야로는 초 해상도(Super-resolution) 기술 등이 있다. 오디오 초 해상도 기술은 저 해상도의 오디오 신호를 기반으로 고 해상도의 오디오를 생성해 내는 기술로 인공 대역 확장(Artificial bandwidth extension), 주파수 대역 확장(Bandwidth extension) 등의 기술 분야로 연구되기도 한다. 주로 연구되는 오디오 대역 확장 기술로는 블라인드 대역 확장 방법이 있으며, 최근에는 딥러닝 기술의 발전에 힘입어 신경망 기반의 오디오 대역 확장 기술들이 많이 연구되고 있다 [01].

그러나 오디오 코덱에서는 고대역 부호화를 위해 블라인드 방식이 아닌 추가 정보를 활용하여 주파수 대역 확장을 수행하는 스펙트럼 대역 복제(SBR: Spectral Band Replication) 기반의 오디오 대역 확장 방법이 주로 연구되어 왔다 [02]. 스펙트럼 대역 복제 기술은 오디오 신호를 코딩하고 전송할 때 낮은 비트율(Bitrate)에서 비트량을 줄이기 위해 인간의 청각 시스템이 저 대역 신호에 비해 고 대역 신호는 상대적으로 둔감하다는 특성을 활용하여, 저 대역 신호는 압축하여 전송하지만 고 대역 신호는 온전히 압축하여 보내지 않고 인코더에서 계산되어 전송된 고주파 신호에 대한 일부 파라미터(Parameters)만을 전송하여 고 대역 신호를 합성하는 코딩 기술이다 [03].

이와 같은 스펙트럼 대역 복제 기반 오디오 대역 확장 기술은 적은 비트만으로도 고 대역 신호를 효과적으로 합성 할 수 있기 때문에 기존 오디오 코덱에서 많이 사용되어 왔다 [04]. 그러나 이러한 방법으로 생성된 고 대역 신호는 복제된 저 대역 신호에 기반하여 고 대역 파라미터를 사용해 신호를 합성하기

때문에, 불 필요한 저 대역 성분이 고 대역 신호에 포함될 수 있다. 따라서 이러한 왜곡 성분을 개선하기 위한 방법들 또한 연구되어 왔으며, 본 논문에서는 이러한 문제점을 해결하고 왜곡을 최소화 하기 위한 방법으로, 잔차 신호 복제 기반의 오디오 대역 확장을 수행하고 부가적으로 복제된 신호를 신경망을 이용해 원본과 유사하도록 품질을 향상시키는 방법을 적용하여 왜곡 성분을 개선하고자 하였다.

2. 제안 방법

본 논문에서는 오디오 코덱에서 활용하는 것을 목적으로 하는 잔차 신호 복제 기반 오디오 대역 확장 방법을 제안하고, 기존 방법과의 성능 비교를 통해 제안 방법의 성능을 검증하였다. 제안하는 방법의 전체 다이어그램(Diagram)은 그림 1 과 같다.

먼저 원본 오디오 신호 x 에서 선형 예측(LP: Linear Prediction)을 수행하여 선형 예측 부호화(LPC: Linear Predictive Coding) 계수(Coefficients)를 추출한다. 다음으로 나머지 잔차 신호(residual)에 대하여 시간/주파수 변환 방법을 이용해 주파수 영역 신호로 변환한다. 이는 오디오 코딩 관점에서 시간/주파수 변환을 수행하면 심리 음향 모델의 분석 및 적용이 용이하기 때문에 코덱에서 주로 사용되는 변환 기법이며, 본 논문에서는 MDCT(Modified Discrete Cosine Transform) 기반의 시간/주파수 변환 방법을 사용했다. 변환된 오디오 신호는 그림 2 와 같이 6.4 kHz 를 기준 대역으로 하여 저 대역 통과 필터를 적용한 뒤 고 대역에 대해서는 저 대역 신호를 이용하여 스펙트럼 대역 복제를 수행 한다. 이와 같은 스펙트럼 대역 복제 방법은 HE-AAC v1 등의 오디오 코덱에서 사용되는 방법으로, 청각이 덜 민감하게 느끼는 고주파 대역의 신호는 실제 양자화하여 전송하지 않고 저 대역 신호와 고 대역 파라미터만을 이용하여 합성하는 코딩 기술이다. 따라서 본 연구에서는 실제 오디오 코덱의 동작 구조를 감안하여 SBR 인코더가 저 대역 신호만을 출력하여 AAC 인코더의 입력으로 전달하고, SBR 디코딩 과정에서는 AAC 디코더에서 저 대역 복원 신호를 입력 받고 여기에 SBR 파라미터를 활용하여 고 대역 신호를 합성해내는 과정을 고려하였다.

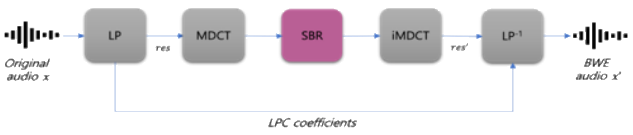


그림 1. 잔차 신호 복제 기반 오디오 대역 확장 방법

Fig. 1. Block diagram of audio bandwidth extension method based on residual signal replication

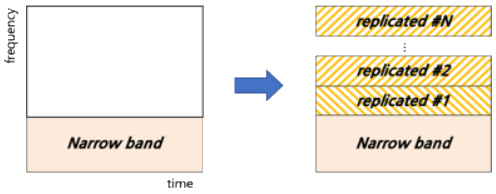


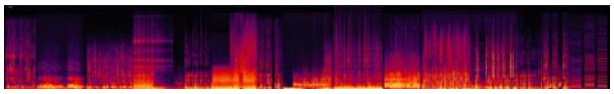
그림 2. 주파수 대역 복제 방법

Fig. 2. Spectral band replication method

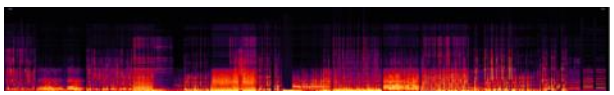
이렇게 복제된 오디오 신호는 전술한 바와 같이 불 필요한 저대역 성분이 고대역에 포함되어 왜곡이 발생하게 된다. 이러한 왜곡 성분은 주파수 대역 별 스케일 팩터(Scale factor) 전송, 잔차 신호 기반 코딩 등을 통해 완화할 수 있지만 완전히 제거하기는 어렵다. 따라서 본 연구에서는 신경망 기반의 오디오 품질 향상 모델을 사용하여 왜곡을 최소화 하고자 하였다. 이를 위해 사용될 수 있는 네트워크 모델은 인공 신경망(Artificial neural network), 오토 인코더(Autoencoder) 등 다양한 모델이 활용될 수 있으며, 본 논문에서는 음성 향상(Speech enhancement) 연구에서 사용되었던 GAN 신경망 모델을 활용하였다 [05].

3. 성능 평가

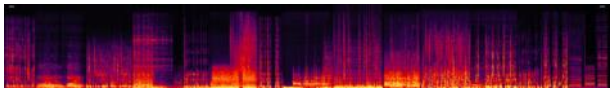
본 절에서는 2 절에서 서술한 대역 확장 방법의 성능을 평가하기 위해 기존 오디오 대역 확장 방법을 베이스라인으로 하여 단계적으로 복원 성능을 측정하였다. 복원 성능에 대한 평가를 위한 지표로는 LSD(Log-spectral distance)가 사용되었으며, 테스트 샘플은 MPEG 에서 오디오 코덱 평가 시 사용되는 테스트 시퀀스 15 개를 활용하였다. 그림 3(a)는 테스트에 사용된 48 kHz 의 표본화율(sampling rate)을 갖는 원본 오디오 샘플의 스펙트로그램이다. 이 원본 신호 대비 6.4 kHz 대역 저주파 통과 필터 적용 신호는 LSD 1.62 dB 의 기준 점수를 보였다.



(a) A spectrogram of the original signal



(b) A bandwidth extension result based on spectral band replication



(c) A spectrogram of the proposed method

그림 3. 제안 방법 실험 결과 스펙트로그램 예시

Fig. 3. Example spectrogram of the proposed method

이를 기준으로 먼저 기존의 오디오 대역 확장 알고리즘의 성능을 비교 검증하기 위해서 신경망 기반의 블라인드 오디오 대역 확장 기술 [06] 및 스플라인 보간법(Spline interpolation)을 이용한 대역 확장 알고리즘의 성능을 평가하였다. 신경망 기반의 오디오 대역 확장 기술은 [01]을 참고하여 동일한 학습 방법 및 데이터를 사용하였으며, 모델 구조는 기준 대역에 맞도록 수정하였다. 즉 6.4 kHz 대역까지는 코어 코딩 기술로 부/복호화 되었음을 가정하고 6.4 kHz 이상의 신호를 대역 확장 기술을 적용하여 복원하였다. 평가 결과 신경망 기반의 블라인드 대역 확장 오디오 신호는 1.57 dB, 스플라인 보간법을 이용한 오디오 대역 확장 신호는 1.47 dB 의 성능을 보였다. 신경망 기반의 대역 확장 방법에서는 테스트 샘플과 다른 종류의 데이터 베이스가 학습에 사용 되었기 때문에 스플라인 보간법을 이용한 대역 확장 기술 대비 낮은 성능을 보인 것으로 판단 된다.

다음으로는 기존 오디오 코덱에서 활용되는 스펙트럼 대역 복제 기반 오디오 대역 확장 기술의 성능을 검토하였다. 주파수 복제 및 합성을 위한 주파수 대역은 5 개의 밴드([6400-7700-9500-12000-15500-24000])로 구분 하였으며, 각 밴드 별 에너지 계수를 추출해 프레임 별 총 5 개의 값을 추가로 전송하여 저 대역 신호로부터 복제된 고 대역 신호의 에너지 값을 스케일링 하는데 사용하였다. 실험 결과 스펙트럼 대역 복제 기반 오디오 대역 확장 기술을 적용한 결과는 그림 3(b)와 같으며, LSD 1.097 dB 로 5 개의 에너지 계수 추가 전송을 통해 저주파 통과 필터 적용 신호 대비 약 0.523 dB 우수한 성능을 얻을 수 있었다.

오디오 부호화에서는 신호에 대한 LPC(Linear predictive coding) 계수를 추출하여 선형 예측을 먼저 수행하면 정보량을 줄일 수 있어 코딩 효율을 높일 수 있다. 또한 주파수 대역 복제 시 발생하는 왜곡 성분을 상대적으로 줄일 수 있기 때문에 본 연구에서는 선형 예측을 통해 정보량을 줄인 뒤 스펙트럼 대역 복제 기반 오디오 대역 확장을 수행하여 더 향상된 오디오 대역 확장 결과를 얻고자 하였다. 선형 예측 부호화 계수는 레빈슨 재귀 알고리즘(Levinson-Durbin recursion algorithm)을 이용하여 추출되었으며, 16 차 LPC 계수를 사용했다. 따라서 매 프레임 별 총 16 개의 추가 계수 전송이 필요하지만, 만약 실제 코어 대역 코딩이 선형 예측 신호 기반으로 수행된다면 이 정보는 추가 비트를 사용하지 않을 수 있다. 실험 결과 선형 예측 및 스펙트럼 대역 복제 기반 오디오 대역 확장 방법은 LSD 0.92 dB 의 성능을 보여, 선형 예측을 수행하지 않은 스펙트럼 대역 복제 기반 오디오 대역 확장 기술 대비 약 0.177 dB 우수한 성능을 얻었다.

마지막으로는 신경망 기반의 오디오 신호 개선 알고리즘을 통해 제안하는 신경망 기반의 오디오 대역 확장 기술의 가능성을 검토하였다. 신경망 기반의 오디오 신호 개선은 [05]의 모델 구조 활용하였으며 왜곡을 최소화 하기 위해 선형 예측을 수행한 잔차 오디오 신호를 사용하여 오디오 신호 개선을 수행하였다. 따라서 입력 신호는 잔차 신호의 대역 복제 신호이며, 출력은 원본 잔차 신호가 된다. 입력 신호로는 한 프레임당 8192 개 샘플의 오디오가 사용되었으며, 50% 오버랩을 통해 배치를 구성하였다. 학습 데이터는 앞서 실험한 신경망 기반 오디오 대역 확장 기술과 동일하게 MedleyDB 를 사용하였으며 [07], 모델 구조는 SEGAN 의 신경망의 레이어 숫자를 간소화 하여 사용하였다 [05]. 최종 실험 결과 테스트 시퀀스의 스펙트로그램은 그림 3(c)와 같으며 LSD 기준 0.84 dB 의 성능을 보여, 선형 예측 및 스펙트럼 대역 복제 기반 오디오 대역 확장 방법 대비 신경망을 이용하여 약 0.08 dB 의 성능 향상을 얻을 수 있었다.

4. 결론

본 논문에서는 선형 예측 기반의 잔차 신호를 이용한 주파수 대역 복제를 통해 추가 파라미터를 활용한 오디오 대역 확장을 수행하고, 대역 복제 시 발생하는 왜곡을 개선하기 위해 신경망 기반 모델의 적용 가능성을 검토하였다. 실험 수행은 MPEG 에서 오디오 코덱 평가 시 사용하는 테스트 시퀀스를 활용하였으며 실험 결과 제안하는 방법을 적용하여 기존 오디오 대역 확장 기술 대비 성능이 향상됨을 확인할 수 있었다. 그러나 선형 예측 및 스펙트럼 대역 복제 기반 방법 대비 신경망을 통한 개선 정도는 크지 않아, 대용량의 데이터베이스를 통한 학습이나 모델 개선 등의 추가적인 연구가 필요할 것으로 판단된다. 또한 객관적 평가 지표 뿐만 아니라 주관적 청취 평가를 통한 성능에 대한 검증도 필요할 것으로 보인다.

감사의 글

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [21ZH1200, 초실감 입체공간 미디어・콘텐츠 원천기술 연구]

참고문헌

[01] 임우택, 백승권, 성종모, 이태진, “신경망 기반 오디오 초해상도 기술 성능 분석,” 한국방송·미디어공학회 하계 학술대회, 2020.

[02] M. Dietz, L. Liljeryd, K. Kjorling and O. Kunz, "Spectral band replication a novel approach in audio coding," Proc. 112th AES Convention, 2002.

[03] 백승권, “디지털 라디오 오디오 코덱 기술의 과거와 미래,” 방송과 기술, v.233, pp.157-163, 2015.

[04] 3GPP TS 26.404: “Enhanced aacPlus encoder SBR part,”

June 2004.

[05] S. Pascual, A. Bonafonte and J. Serra. “SEGAN: Speech enhancement generative adversarial network,” arXiv preprint arXiv:1703.09452, 2017.

[06] V. Kuleshov, S. Z. Enam and S. Ermon, "Audio Super Resolution using Neural Nets", International Conference on Learning Representation (ICLR), 2017.

[07] R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam and J. Bello, “MedleyDB: A Multitrack dataset for annotation-intensive MIR research”, 15th International Society for Music Information Retrieval Conference (ISMIR), 2014.