특집논문 (Special Paper)
방송공학회논문지 제24권 제3호, 2019년 5월 (JBE Vol. 24, No. 3, May 2019)
https://doi.org/10.5909/JBE.2019.24.3.387
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

부가 정보를 이용하는 오토 인코더 기반의 오디오 고대역 부호화 기술

조 효 진[a], 신 성 현[a], 백 승 권[b], 이 태 진[b], 박 호 종[a]‡

Audio High-Band Coding based on Autoencoder with Side Information

Hyo-Jin Cho[a], Seong-Hyeon Shin[a], Seung Kwon Beack[b], Taejin Lee[b], and Hochong Park[a]‡

요 약

본 논문에서는 부가 정보를 이용하는 오토 인코더 기반의 새로운 오디오 고대역 부호화 방법을 제안한다. 제안하는 방법은 MDCT 영역에서 동작하고, 부호화 할 정보만 입력하는 기존의 오토 인코더와 다르게, 과거와 현재의 저대역 정보로 구성된 부가 정보를 추가로 입력하여 오토 인코더의 복원 성능을 향상시킨다. 특히, 시간-주파수 영역의 부가 정보를 사용하여 시간에 따른 신호 특성을 고대역 복원에 활용하도록 한다. 제안하는 방법에서 부호화기는 매 프레임마다 오토 인코더가 생성한 4차원 latent 벡터와 이득 정보를 12비트로 양자화 하여 전송한다. 복호화기는 과거와 현재 프레임에서 복원된 저대역 정보와 전송 받은 정보를 오토 인코더에 입력하여 고대역 정보를 복원한다. 청취 평가를 통하여 제안하는 방법이 SBR에 비하여 약 1/2의 비트율로 SBR과 동등 품질의 고대역 정보를 복원하는 것을 확인하였다.

Abstract

In this study, a new method of audio high-band coding based on autoencoder with side information is proposed. The proposed method operates in the MDCT domain, and improves the performance by using additional side information consisting of the previous and current low bands, which is different from the conventional autoencoder that only inputs information to be encoded. Moreover, the side information in a time-frequency domain enables the high-band coder to utilize temporal characteristics of the signal. In the proposed method, the encoder transmits a 4-dimensional latent vector computed by the autoencoder and a gain variable using 12 bits for each frame. The decoder reconstructs the high band by applying the decoded low bands in the previous and current frames and the transmitted information to the autoencoder. Subjective evaluation confirms that the proposed method provides equivalent performance to the SBR at approximately half the bit rate of the SBR.

Keyword : autoencoder, neural network, audio high-band coding, side information

a) 광운대학교 전자공학과(Dept. of Electronics Engineering, Kwangwoon University)
b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)
‡ Corresponding Author : 박호종(Hochong Park)
E-mail: hcpark@kw.ac.kr
Tel: +82-2-940-5104
ORCID: https://orcid.org/0000-0003-1600-6610

※ 본 논문은 2018년도 광운대학교 교내학술연구비 지원과 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발).
※ The present Research has been conducted by the Research Grant of Kwangwoon University in 2018 and by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2017-0-00072002, Development of audio/video coding and light field media fundamental technologies for ultra realistic tera-media).
· Manuscript received March 15, 2019; Revised April 30, 2019; Accepted April 30, 2019.

Ⅰ. 서 론

고전적인 오디오 부호화는 주파수 영역에서 주파수 계수를 양자화 하는 변환 부호화 방법을 사용하고, 심리음향 모델을 기반으로 양자화 잡음을 제어하여 주어진 비트율에서 음질 왜곡을 최소화 한다[1]. 변환 부호화는 높은 비트율에서 우수한 오디오 품질을 제공하지만, 비트율이 낮아지면 주파수 정보의 왜곡이 심해져 음질이 저하되는 문제점을 가진다.

낮은 비트율에서 변환 부호화보다 우수한 성능을 가지는 새로운 오디오 부호화 기술 연구가 진행되었고, 오디오 정보를 파라미터 영역에서 표현하여 부호화 하는 파라메트릭 부호화 기술이 개발되었다[2]. 파라메트릭 부호화는 전송할 오디오 정보를 소수의 파라미터로 표현하여 전송하므로 비트율을 낮출 수 있지만, 원 정보의 손실이 크게 발생하여 고품질의 부호화를 수행하지는 못한다. 따라서 음질에 큰 영향을 주는 오디오 저대역은 변환 부호화로 전송하고, 청각적으로 중요도가 낮은 고대역은 파라메트릭 부호화로 전송하는 방법이 널리 사용된다. 대표적인 고대역 파라메트릭 부호화 방법에 spectral band replication (SBR)이 있다[2]. SBR 부호화기는 오디오 신호를 quadrature mirror filter (QMF)를 사용하여 시간-주파수 영역으로 변환하고, 고대역을 블록 단위로 나누고 각 블록 단위로 에너지와 톤 정보 (tonality) 등의 파라미터를 구하여 전송한다. 복호화기는 시간-주파수 영역에서 저대역 정보를 고대역으로 복사하고, 전송된 고대역 파라미터를 적용하여 최종 고대역 신호를 생성한다. SBR을 사용하여 오디오 부호화기를 동작시킬 경우, 변환 부호화하는 저대역의 동작 영역이 고대역의 QMF 영역과 다르므로 부호화 과정에서 두 종류의 변환을 수행하여 계산량이 증가하는 문제점이 나타나고 따라서 QMF를 사용하지 않는 고대역 파라메트릭 부호화 방법이 요구된다[3].

최근 신경망을 이용하여 고대역을 복원하는 방법이 널리 연구되고 있다[4,5]. 고대역 신호를 여기 신호와 스펙트럼 포락선으로 모델링 하고, 순환 신경망 (recurrent neural network)과 컨벌루션 신경망 (convolutional neural network, CNN)을 이용하여 과거와 현재 프레임의 저대역 정보로부터 고대역 여기 신호 또는 포락선을 예측한다. 특히, 긴 시간의 과거 저대역 정보를 활용하여 고대역 정보를 예측함으로써 성능을 향상시킨다.

본 논문에서는 오토 인코더 (autoencoder)를 이용하는 새로운 고대역 파라메트릭 부호화 방법을 제안한다[6]. 부호화할 정보만을 입력하는 기존의 오토 인코더 동작과 다르게, 제안하는 오토 인코더는 과거와 현재의 저대역 정보로 구성된 부가 정보를 입력하여 복원 성능을 향상시킨다. 이 구조를 사용하면 시간-주파수 영역의 부가 정보를 사용하여 시간에 따른 신호 특성을 고대역 복원에 활용할 수 있다. 제안한 방법은 MDCT (modified discrete cosine transform) 영역에서 진행되며, SBR에서 필요한 QMF 계산을 제거하여 계산량 감소를 얻고 저대역과 고대역을 모두 MDCT 영역에서 수행하는 장점을 가진다. 또한, 제안한 방법은 기존 신경망 기반의 부호화 방법에서 사용하는 고대역 신호의 생성 모델을 사용하지 않고 고대역의 MDCT 계수를 직접 예측하는 차별성을 가진다. 제안하는 오토 인코더 기반의 고대역 부호화 성능을 SBR 성능과 비교하였고, 청취 평가를 통하여 제안 방법이 SBR보다 약 1/2의 비트율로 동등한 품질의 고대역 정보를 복원하는 것을 확인하였다.

Ⅱ. 제안하는 고대역 부호화 방법

1. 오토 인코더 개요

그림 1이 오토 인코더의 기본 구조를 보여준다. 오토 인코더의 한 은닉층 (hidden layer) 차원을 입력층 차원보다 매우 작게 하고 입력과 출력이 동일하도록 훈련 하면, 오토 인코더는 입력 정보를 해당 은닉층의 적은 데이터로 압축

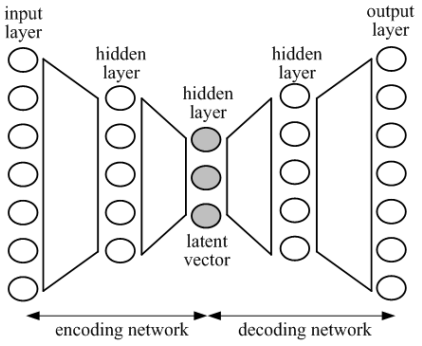


그림 1. 오토 인코더의 기본 구조
Fig. 1. Basic structure of autoencoder

하여 표현하게 된다. 이와 같은 동작에서 입력 정보를 압축하여 표현한 은닉층 값을 특별히 latent 벡터라 한다. 또한, 그 이후의 신경망 동작을 통하여 압축 정보로부터 다시 입력 신호로 복원할 수 있다. 즉, 부호화 동작은 오토 인코더의 인코딩 신경망 (encoding network)을 이용하여 입력 신호에 대한 latent 벡터를 구하는 과정이고, 복호화 동작은 디코딩 신경망 (decoding network)을 이용하여 latent 벡터로부터 원 신호로 복원하는 과정이다.

2. 제안하는 오토 인코더 구조

제안하는 부호화기는 1024 샘플 길이의 프레임 단위로 동작하고, 50% 중첩을 가지는 2048 샘플 길이의 윈도우를 적용하고 MDCT를 계산하여 각 프레임별로 1024개 MDCT 계수를 구한다. 부호화 대역폭은 14.25 kHz로 정하고, 9.75 ~ 14.25 kHz을 고대역이라 정의한다. 샘플링 주파수는 48 kHz이고, 한 프레임은 총 608개 MDCT 계수를 가지고 고대역은 192개 MDCT 계수로 구성된다.

그림 2가 본 논문에서 제안하는 오토 인코더의 전체 구조를 보여준다. 그림 1과 같이 부호화 할 정보만을 입력하는 기존 오토 인코더 구조와 다르게, 부가 정보를 추가로 이용하는 구조를 가지고 두 개의 인코딩 신경망이 병렬 구조로 존재한다. 부호화할 정보에 해당하는 192개 고대역 MDCT 계수가 첫 번째 인코딩 신경망에 입력되고, 총 3개 층의 FCN (fully-connected network)를 거쳐 4차원 latent 벡터 $\boldsymbol{X}$ 로 변환된다[7]. 부가 정보는 현재와 7개 과거 프레임의 3.75 ~ 9.75 kHz 영역의 MDCT 계수로 구성되고, 8 × 256 구조의 2차원 (2D) 데이터이다. 3.75 kHz 이하 정보는 고대역과 상관관계가 낮으므로 부가 정보로 사용하지 않는다. 이전 프레임 정보를 부가적으로 사용하여 신호의 시간적 변화 정보를 고대역 복원에 활용하여 오토 인코더의 복원 성능을 향상시킬 수 있다. 부가 정보는 두 번째 인코딩 신경망에 입력되어 총 3개 층의 2D CNN과 1차원 평탄화 (flatten)와 FCN을 통과하여 10차 latent 벡터 $\boldsymbol{Y}$ 로 변환된다[7]. 2D CNN을 사용함으로써 부가 정보의 주파수 특성뿐만 아니라 시간적 특성을 활용할 수 있게 하였다. 이와 같이 구해진 두 종류의 latent 벡터 $\boldsymbol{X}$ 와 $\boldsymbol{Y}$ 를 결합하여 14차 latent 벡터를 얻고, 이를 디코딩 신경망에 입력하여 최종 출력 데이터를 구한다. 오토 인코더의 상세 신경망 구조는 표 1에 정리되어 있다.

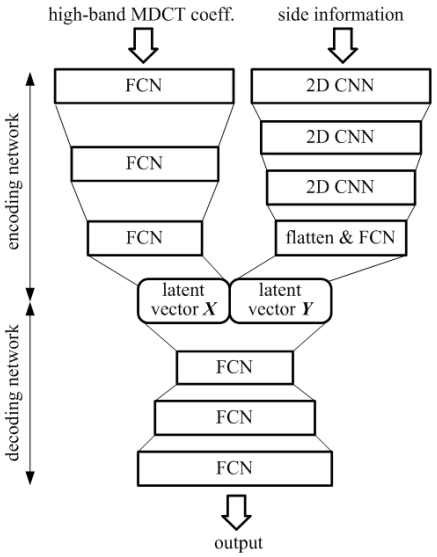


그림 2. 제안하는 오토 인코더 구조
Fig. 2. Structure of the proposed autoencoder

표 1. 제안하는 방법에서 사용하는 신경망의 세부 구조
Table 1. Detail of network structure in the proposed method

Encoding network for high band

layer	function	output dim.
in	high-band MDCT coeff.	192
1	FCN, GLU	96
2	FCN, GLU	24
3	FCN, sigmoid	4
out	latent vector	4

Encoding network for side information

layer	function	output dim.	filters	kernel	stride
in	side-info. MDCT coeff.	8×256			
1	2D CNN, GLU	4×128×32	32	[5,5]	[2,2]
2	2D CNN, GLU	2×64×64	64	[5,5]	[2,2]
3	2D CNN, GLU	1×32×128	128	[5,5]	[2,2]
4	flatten, FCN, sigmoid	10	-	-	-
out	latent vector	10			

Decoding network

layer	function	output dim.
in	latent vector	14
1	FCN, GLU	32
2	FCN, GLU	96
3	FCN, sigmoid	192
out	high-band MDCT coeff.	192

부가 정보를 압축하지 않고 그대로 latent 벡터로 사용하면 부가 정보의 정보량이 매우 크게 되고, 디코딩 신경망이 고대역을 복원할 때 복원의 목표가 되는 입력 정보 보다 부가 정보에 더 의존하게 되어 고대역 복원 성능이 저하되는 문제를 가진다. 또한, 너무 적은 데이터로 압축하면 고대역 복원에 활용할 정보가 부족하여 성능 향상에 한계를 가진다. 본 논문에서는 다양한 크기의 $\boldsymbol{Y}$ 에 대한 실험을 진행하고 성능을 분석하여 부가 정보를 10개 데이터로 압축하도록 최종 결정하였다.

인코딩 신경망과 디코딩 신경망의 출력단 활성화 함수로 sigmoid를 사용하고, 은닉층은 그림 3과 같은 GLU (gated linear unit)을 사용한다[8]. 이전 층의 출력 h_{t-1} 에 가중치 W 를 곱하고 바이어스 b 를 더한 결과인 z 를 구하고, z 에 tanh와 sigmoid 함수를 각각 적용하고 곱하여 현재 층의 GLU 출력 h_t 를 구한다. GLU를 사용하면 다음 층에 전달하는 tanh 출력의 비율을 sigmoid의 출력을 통해 조절할 수 있어 더 다양한 신경망 동작을 수행할 수 있다.

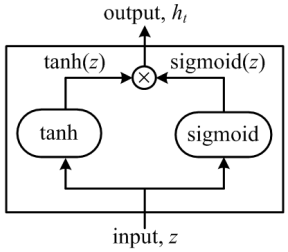


그림 3. GLU 구조
Fig. 3. GLU structure

그림 2의 오토 인코더 훈련은 다음과 같이 진행한다. 훈련 데이터로부터 192개 고대역 MDCT 계수와 8 × 256 크기의 부가 정보를 생성하여 각각 인코딩 신경망에 입력하고, 디코딩 신경망의 최종 출력 192개 데이터를 구한다. 다음, 입력 고대역 MDCT 계수와 출력 사이의 오차가 최소가 되도록 전체 신경망을 동시에 훈련한다. 손실 함수로 입력과 출력 사이의 최소 평균 제곱 오차를 사용하였고 신경망 훈련은 ADAM을 사용하였다[9]. 실제 복호화기를 동작시킬 때, latent 벡터 $\boldsymbol{X}$ 가 양자화 오차를 포함하고 부가 정보는 변환 부호화에 의한 저대역 양자화 오차를 포함하므로 훈련에 사용하였던 정보와 일치하지 않고, 이에 따라 복호화기의 성능이 저하될 수 있다. 그러나 본 논문에서는 훈련의 독립성과 일반성을 확보하기 위해 양자화 하지 않은 $\boldsymbol{X}$ 와 부가 정보를 사용하여 신경망을 훈련한다.

3. 양자화

4차원 latent 벡터 $\boldsymbol{X}$ 의 양자화기 설계를 위해 $\boldsymbol{X}$ 의 특성을 분석하였다. 그림 4가 $\boldsymbol{X}$ 의 각 요소값 사이의 2차원 분포도를 보여주며, x축과 y축은 각 값을 32개 영역으로 균등하게 나눌 때의 영역 인덱스이다. 각 쌍 사이에 큰 상관관계가 있음을 알 수 있고, 따라서 벡터 양자화를 통하여 양자화 성능을 향상시킬 수 있다. 본 논문에서는 훈련 데이터를 사용하여 k-평균 알고리즘 기반으로 8-비트 벡터 양자화기를 설계하였다.

제안 방법에서 설계한 8-비트 벡터 양자화기의 성능을 검증하기 위해 8-비트 균등 양자화기의 성능과 비교하였고, 제안한 벡터 양자화기가 균등 양자화기에 비해 양자화 오차의 제곱 평균이 약 1/6이 되는 것을 확인했다. 또한, 균등 양자화기가 제안한 양자화기와 동등한 성능을 가지려면 12비트 이상이 필요한 것을 확인하였다.

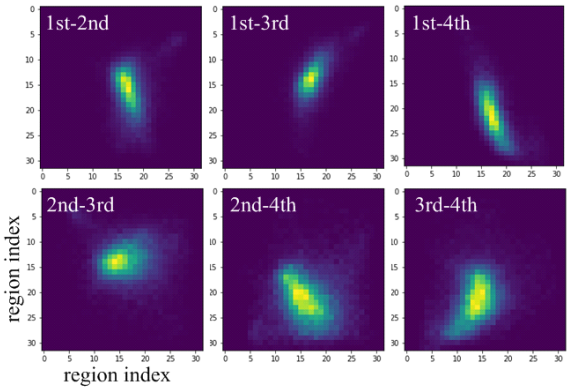


그림 4. Latent 벡터 $\boldsymbol{X}$ 값의 2차원 분포도
Fig. 4. 2D scatter diagram of latent vector $\boldsymbol{X}$

4. 고대역 부호화 및 복호화 동작

앞에서 설명하였듯이 제안한 방법은 MDCT 영역에서 수행되는데, MDCT 계수의 부호 (sign)는 불규칙 성질이 매우 강하여 MDCT 계수로부터 핵심 정보를 추출하는 것을 방해하고 저대역과 고대역 사이의 상관관계를 저하시킨다. 즉,

MDCT 계수를 사용하여 그림 2의 오토 인코더를 훈련시키면 원하는 부호화기 성능을 얻기 어렵다. 이를 해결하기 위하여 제안하는 방법은 MDCT 계수 크기를 오토 인코더에 입력하고 MDCT 계수 크기를 복원하도록 하고, MDCT 계수 부호는 별도로 처리한다. MDCT 계수의 부호를 그대로 사용할 때와 크기만을 사용할 때에 오토 인코더의 고대역 복원 성능에 큰 차이가 있음을 실험을 통해 확인하였다.

오디오 신호의 프레임별 MDCT 계수 크기에 많은 차이가 있으므로 MDCT 크기를 정규화 하여 신경망에 입력하여야 훈련 성능이 향상되고 궁극적으로 부호화 성능이 향상된다. 따라서 MDCT 계수 크기를 0 ~ 1 사이로 정규화한 후 오토 인코더에 입력하고 훈련시킨다. 정규화는 각 프레임 별로 고대역 MDCT 계수 크기 합과 저대역 MDCT 계수 크기의 합이 각각 1이 되도록 하였다. 이와 같은 프레임별 정규화를 통하여 프레임 에너지 차이에 의한 오토 인코더 성능 변화를 줄일 수 있다. 정규화의 복원을 위해 정규화에 사용하였던 고대역 MDCT 크기 합을 이득 변수 G 로 정의하고, 이 값을 양자화 하여 복호화기로 전송한다. 복호화기는 오토 인코더가 출력하는 MDCT 계수 크기에 전송된 G 를 곱하여 최종 MDCT 계수 크기를 얻는다. 이득 변수 G 값은 4 비트로 스칼라 양자화 하고, k-평균 알고리즘을 사용하여 양자화기를 설계하였다.

정규화 된 MDCT 계수 크기를 사용하여 그림 2의 오토 인코더를 훈련하고, 훈련된 오토 인코더를 이용하여 다음과 같이 고대역 부호화를 실시한다. 먼저, 입력 오디오 신호에서 부호화 할 고대역 MDCT 계수 192개를 구하여 정규화 하고, 이득 변수 G 를 구하고 4 비트로 양자화 하여 전송한다. 다음, 192개의 정규화 된 MDCT 계수 크기를 인코딩 신경망에 입력하여 4차원 벡터 $\boldsymbol{X}$ 를 구하고, 8 비트로 벡터 양자화 하여 전송한다. 즉, 총 전송하는 데이터는 5개이고, 양자화 비트는 12개이고, 따라서 고대역 부호화의 비트율은 0.56 kbps이다.

고대역 복호화 과정은 다음과 같다. 현재와 7개 과거 프레임의 저대역 MDCT 계수 크기를 구하고 정규화 한 후 인코딩 신경망에 입력하여 10차 벡터 $\boldsymbol{Y}$ 를 구한다. 다음, $\boldsymbol{Y}$ 와 전송된 4차 벡터 $\boldsymbol{X}$ 를 결합하여 14차 벡터를 구하고 디코딩 신경망에 입력하여 출력을 구하고, 전송된 이득 변수 G를 출력에 곱하여 고대역 MDCT 계수 크기를 구한다. 마지막으로, intelligent gap filling (IGF)과 유사하게 저대역 MDCT 계수 부호를 고대역 MDCT 계수 크기에 적용하여 최종 고대역 MDCT 계수를 구한다[3]. 이와 같은 부호 복사를 통해 이웃한 MDCT 계수 부호의 연결 패턴을 재활용 할 수 있고, 무작위로 MDCT 부호를 할당하는 것에 비하여 고대역 복원 성능이 향상되는 것을 확인하였다.

Ⅲ. 성능 평가

신경망 훈련을 위한 훈련 데이터는 VCTK (voice cloning toolkit) 데이터 세트[10], RWC (real world computing) 데이터 세트[11], 베토벤 피아노 소나타 데이터 세트이고, 총 길이는 약 57 시간이다. 평가 데이터는 USAC (unified speech and audio coding) 평가에 사용 하였던 12개 클립이고, 각 4개씩의 클립을 가지는 ‘speech’, ‘speech-over-music’ (SoM), ‘music’ 의 3개 그룹으로 구성된다[12].

제안한 고대역 부호화는 저대역 MDCT 계수를 부가 정보로 사용하므로 저대역 MDCT 계수의 양자화를 포함하여 성능을 평가해야 한다. 저대역 부호화는 48 kbps USAC을 사용하여 실시한다[13]. 즉, 독립적으로 48 kbps USAC을 실행하여 MDCT 계수를 양자화 하고, 양자화된 MDCT 계수를 9.75 kHz 이하의 저대역에 적용한 후에 제안한 고대역 부호화기를 동작시킨다. 이렇게 하면 저대역은 USAC으로 부호화 하고 고대역은 제안 방법으로 부호화 한 출력 신호를 얻을 수 있고, 이 신호에 대한 청취 평가를 진행한다. 성능 평가는 모노 신호에 대하여만 진행하고, 모든 프레임은 long window로 처리하도록 하였다. 멀티채널과 short window 성능은 추후에 보고할 예정이다.

제안한 고대역 부호화의 성능을 SBR 성능과 비교 하였다. SBR을 적용할 대역은 자체 규격에 따라 결정되는데, 제안 방법의 고대역과 가장 근사하게 설정할 경우 10.125 kHz 이상이 된다[13]. 따라서 고대역을 10.125 ~ 14.25kHz로 설정하고 SBR를 동작시켜 고대역을 복원한다. 이 경우 평가 데이터에 대한 SBR 비트율은 평균 1.08 kbps이고, 제안 방법의 약 2배의 평균 비트율을 가진다. 저대역은 제안 방법과 동일하게 48 kbps USAC에 따라 부호화 하여 사용한다. 즉, 제안한 방법과 SBR에 대한 평가 오디오 신호에서, 저대역

은 동등한 방법으로 부호화 되고 고대역만 서로 다른 방법으로 부호화 된다. 따라서 두 오디오 신호의 품질을 비교하면 고대역 부호화에 의한 성능 차이만 평가할 수 있다.

그림 5는 원본 신호, 제안하는 방법과 SBR을 사용해 복원한 신호의 스펙트로그램을 보여준다. 제안한 방법은 SBR에 비해 약 1/2의 비트만 사용하지만 SBR과 동등 이상의 품질을 복원하는 것을 확인할 수 있다. 주관적 청취 평가는 MUSHRA 방법으로 진행하였고, 앵커로 3.5 kHz 저대역 통과 신호를 사용하였다[14]. 청취 평가는 5명이 진행하였고, 원본, 두 가지 방법으로 복원한 신호, 앵커 신호를 자유롭게 선택하여 청취하고, 각각의 청취 품질을 0 ~ 100점 사이의 점수로 나타낸다. 그림 6이 각 평가 데이터 그룹과 전체 평균 결과를 보여주며, 95% 신뢰구간을 나타내었다. 저대역은 동일한 방법으로 부호화 하였으므로 청취 품질의 차이는 고대역의 품질 차이에 의한 것이다. 평균적으로 제안한 방법과 SBR 방법의 청취 품질이 동등하고, 'speech'에 대하여 제안 방법이 SBR 방법보다 높은 품질을 제공한다. 따라서 제안한 방법이 SBR에 비하여 고대역 비트 수를 1/2만 사용하고 동등한 청취 품질을 제공하는 것을 알 수 있다.

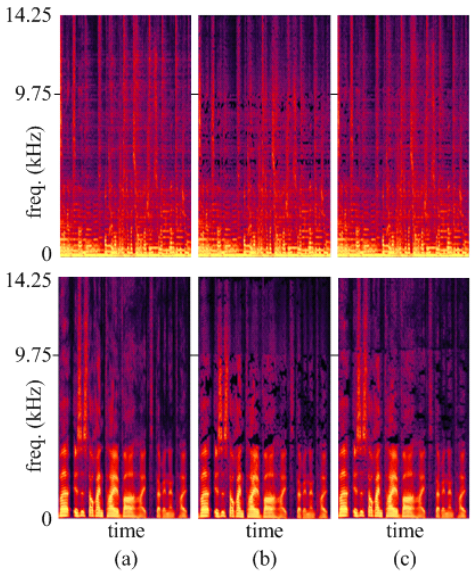


그림 5. 평가 데이터의 스펙트로그램
(a) 원본 신호, (b) 제안 방법으로 복원한 신호, (c) SBR로 복원한 신호
Fig. 5. Spectrogram of test data
(a) original, (b) decoded signal by proposed method and (c) decoded signal by SBR

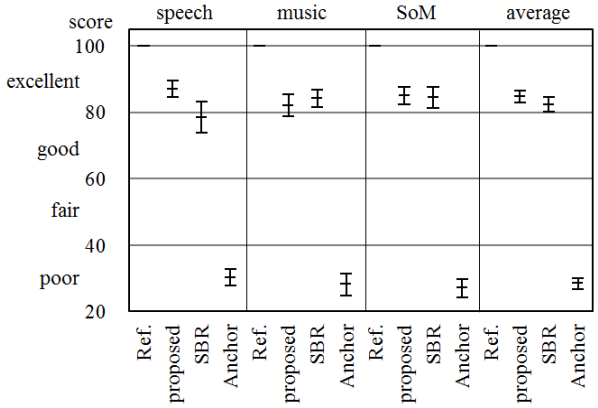


그림 6. MUSHRA 청취 평가 결과
Fig. 6. Result of MUSHRA test

Ⅳ. 결 론

본 논문에서는 부가 정보를 이용하는 오토 인코더를 사용하여 오디오 고대역을 부호화 하는 방법을 제안하였다. 부호화할 정보뿐만 아니라 현재와 과거 정보로 구성된 부가 정보를 추가로 입력하여 과거 정보와 저대역 정보를 동시에 활용할 수 있게 하였다. 또한, 부가 정보를 인코딩 신경망을 통하여 압축 데이터로 변환 한 후 디코딩 신경망에 입력하여 복원 성능을 향상시켰다. 제안한 부호화 방법은 고대역 정보를 4개의 압축 데이터와 한 개의 이득 정보로 표현하고, 0.56 kbps 비트율을 가진다. 청취 평가를 통하여 제안 방법이 SBR보다 약 1/2 비트율로 동등한 청취 품질을 제공하는 것을 확인하였다. 추가 연구를 통하여 고대역의 대역폭을 증가시키고 낮은 대역에서의 복원 성능을 높이기 위한 새로운 복원 방법과 네트워크 구조에 대하여 연구하고 생성 모델 기반으로 고대역 부호화의 성능을 향상시키는 방법을 연구할 계획이다.

참 고 문 헌 (References)

[1] ISO/IEC 11172-3, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3," 1993.
[2] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral band replication, a novel approach in audio coding," *112th Conv. Audio Eng. Soc.*, May 2002.
[3] C. R. Helmrich, et al., "Spectral envelope reconstruction via IGF for

audio transform coding," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Brisbane, Australia, pp. 389-393, 2015.
[4] L. Jiang, R. Hu, X. Wang, W. Tu, and M. Zhang, "Nonlinear prediction with deep recurrent neural networks for non-blind audio bandwidth extension," *China Communication*, vol. 15, no. 1, pp. 72-85. Jan. 2018.
[5] K. Schmidt and B. Edler, "Blind bandwidth extension based on convolutional and recurrent deep neural networks," *Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Calgary, Canada, pp. 5444-5448, 2018.
[6] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, 313.5786, pp. 504-507, 2006.
[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521.7553, pp. 436-444, 2015.
[8] Y. N. Dauphin, et al., "Language modeling with gated convolutional networks," *Proc. of the 34th Int. Conf. on Machine Learning*, vol 70, Sydney, Australia, pp. 933-941, 2017.
[9] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *Proc. of Int. Conf. on Learning Representation*, San Diego, USA, 2015.
[10] C. Veaux, et al., "Superseded-CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," 2016.
[11] M. Goto, "Development of the RWC music database," *Proc. of Int. Congress on Acoustics*, vol. 1, pp. 553-556, April 2004.
[12] ISO/IEC JTC1/SC29/WG11 N9927, "Workplan for subjective testing of Unified Speech and Audio Coding proposals," April 2008.
[13] S. Beack, et al., "Single-mode-based Unified Speech and Audio Coding by extending the linear prediction domain coding mode," *ETRI Journal*, vol. 39, no. 3, pp. 310-318, 2017.
[14] ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," 2015.

저 자 소 개

조 효 진

- 2017년 2월 : 광운대학교 전자공학과 공학사
- 2017년 3월 ~ 현재 : 광운대학교 전자공학과 석사과정
- ORCID : http://orcid.org/0000-0003-2296-2270
- 주관심분야 : 오디오/음성 신호처리, 딥 러닝

신 성 현

- 2016년 2월 : 광운대학교 전자공학과 공학사
- 2016년 3월 ~ 현재 : 광운대학교 전자공학과 석박사통합과정
- ORCID : http://orcid.org/0000-0002-2343-8983
- 주관심분야 : 오디오/음성 신호처리, 딥 러닝

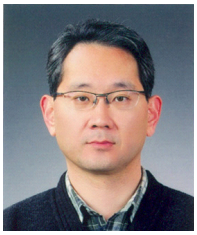
백 승 권

- 2005년 8월 : 한국과학기술원 통신공학부 공학박사
- 2005년 8월 ~ 현재: 한국전자통신연구원 실감AV연구그룹 책임연구원
- ORCID : https://orcid.org/0000-0002-6254-2062
- 주관심분야 : 오디오/음성 신호처리

저 자 소 개

이 태 진

- 2014년 : 충남대학교 전자전파정보통신공학과 공학박사
- 2002년 ~ 2003년 : 일본 Tokyo Denki University, 방문연구원
- 2000년 ~ 현재 : ETRI 실감AV연구그룹 책임연구원
- 주관심분야 : 오디오 부호화, 실감음향, 오디오 신호처리



박 호 종

- 1986년 2월 : 서울대학교 전자공학과 공학사
- 1987년 12월 : Univ. of Wisconsin-Madison 공학석사
- 1993년 5월 : Univ. of Wisconsin-Madison 공학박사
- 1993년 9월 ~ 1997년 8월 : 삼성전자 선임연구원
- 1997년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- ORCID : https://orcid.org/0000-0003-1600-6610
- 주관심분야 : 오디오/음성 신호처리, 3D 오디오, 음악정보처리