

레터논문 (Letter Paper)

방송공학회논문지 제26권 제3호, 2021년 5월 (JBE Vol. 26, No. 3, May 2021)

<https://doi.org/10.5909/JBE.2021.26.3.327>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

수정된 잡음 채움 기법을 적용한 3D 오디오 부호기

김인철^{a)†}

A 3D Audio Codec Employing a Revised Noise Filling Method

Rin Chul Kim^{a)†}

요 약

본 논문에서는 3D 오디오 부호기의 성능 향상을 위해 새로운 잡음 채움 기법을 제안한다. 제안한 잡음 채움 기법에서는 핵심 대역을 IGF 시작 주파수까지가 아니라 MAX_SFB로 설정되는 영역까지로 제한한다. 그리고, IGF 입력 패치의 모든 주파수 영역에 잡음 채움이 수행되도록 한다. 본 논문에서 수정된 잡음 채움 기법의 성능은 MUSHRA 테스트를 통해 평가되었고, 그 결과 기존의 방법보다 우수한 성능을 보임을 입증한다.

Abstract

In this paper, a new noise filling method is proposed for improving the performance of the 3D audio codec. In the new method, the core band is limited up to MAX_SFB, not up to the IGF start frequency. And the noise filling is applied to all frequency range of the IGF source patches. We conduct the MUSHRA test and find that the proposed noise filling method demonstrates better performance than the conventional method.

Keyword : 3D audio coding, noise filling, intelligent gap filling

1. 서 론

MPEG 표준 초기부터 오디오 부호기에 도입되었던

a) 서울시립대학교 전자전기컴퓨터공학부(Dept. of ECE, Univ. of Seoul)

† Corresponding Author : 김인철(Rin Chul Kim)

E-mail: rin@uos.ac.kr

Tel: +82-2-6490-2336

ORCID: <https://orcid.org/0000-0002-7643-8717>

※ This work was supported by the 2018 sabbatical year research grant of the University of Seoul.

· Manuscript received March 5, 2021; Revised April 9, 2021; Accepted April 9, 2021.

MDCT (modified discrete cosine transform) 기반 부호기는 그 성능이 향상되어 가장 최신의 MPEG 오디오 표준인 3DA (3D audio)^[1]에서 FD (frequency domain) 모드라는 이름으로 포함되어 있다. FD 모드의 3DA 부호기는 매우 우수한 성능을 보이고 있지만, 저 전송률에서는 거친 양자기 때문에 주관적 품질이 크게 낮아진다는 단점이 있다. 특히, MDCT 계수가 0으로 양자화되어 해당 주파수 성분이 없어지는 스펙트럼 공백 (spectral hole)은 귀에 매우 거슬리는 왜곡 현상을 유발할 수 있다.

거친 양자화기에 의한 품질 저하는 심리음향 모델에 따른 효율적인 비트 할당과 함께 잡음 채움 (noise filling)^[2,3] 과 IGF (intelligent gap filling)^[4] 등을 통해 효과적으로 개선할 수 있다. IGF는 기존의 SBR (spectral band replication)을 대체할 수 있는 기법으로, MDCT 영역에서 대역 확장을 수행하는 새로운 기법이다. IGF는 사람 귀에 덜 민감한 고주파 영역을 작은 비트량의 파라미터만으로 표현함으로써, 민감한 저주파 영역을 많은 비트량으로 충실히 부호화할 수 있게 한다. 여기서 저주파 영역 부호화에서 발생하는 스펙트럼 공백은 복호기에서 잡음 채움을 통해 원 신호와 동등한 에너지 레벨의 잡음으로 복원하여 주관적 음질을 개선시킬 수 있도록 한다.

본 논문에서는 IGF와 잡음 채움의 효과적 결합에 대해 살펴보고자 한다. 3DA 표준^[1]에서는 IGF 시작 주파수까지의 저주파 영역은 MDCT 기반 부호기 및 잡음 채움으로 부호화한다. 그리고, 고주파 영역은 복원된 저주파 영역과 함께 작은 양의 IGF 파라미터들을 이용하여 복원한다. 이러한 방법은 매우 낮은 전송률에서는 주관적 품질이 크게 저하될 수 있는 단점이 있다. 본 논문에서는 IGF를 채택할 때 효과적인 잡음 채움 기법을 제안하겠다. 제안한 기법에서는 충실히 부호화되는 저주파 영역을 IGF 시작 주파수까지가 아니라 MAX_SFB로 설정되는 영역까지 제한한다. 그리고, IGF 입력 패치는 모두 잡음 채움을 수행하도록 한다. 제안한 잡음 채움 기법의 성능은 복원된 오디오 신호에 대해 주관적 음질 평가를 시행함으로써 평가하겠다.

II. 잡음 채움

FD 모드로 부호화할 때 전송률이 낮아지면 매우 거친 양자화로 인해 많은 MDCT 계수들이 0으로 표현된다. 이때 0으로 표현된 MDCT 계수들 때문에 날카롭거나 불안정한 소리로 복원될 수 있다^[3]. 3DA^[1]에서는 3.5KHz 이상에서 0으로 양자화된 MDCT 계수들을 작은 값으로 대체함으로써 주관적 품질을 개선한다. 이러한 기법을 잡음 채움^[1,3]이라고 하는데 좀 더 구체적으로 설명하면 다음과 같다.

SF (scale factor)가 $scf[b]$ 인 b 번째 SFB (scale factor band)에 속한 n 번째 MDCT 계수의 양자화된 값 $q[n]$ 은 복호기에서 다음과 같이 MDCT 계수 $\hat{x}[n]$ 으로 역양자화된

다.

$$\hat{x}[n] = \pm |q[n]|^{4/3} \times 2^{0.25scf[b]-25} \quad (1)$$

$q[n]$ 값이 0이라면 $\hat{x}[n]$ 은 0으로 복원된다. 그러나 잡음 채움 기법을 적용한다면 $\hat{x}[n]$ 은 0이 아니라 잡음 레벨 (noise_level) 값을 이용하여 다음과 같이 복원된다.

$$\hat{x}[n] = \pm 2^{(noise_level-14)/3} \times 2^{0.25scf[b]-25} \quad (2)$$

만약 SFB 내의 모든 MDCT 계수가 0이라고 하면, $\hat{x}[n]$ 은 잡음 레벨과 함께 잡음 오프셋 (noise_offset) 값을 이용하여 다음과 같이 복원된다.

$$\hat{x}[n] = \pm 2^{(noise_level-14)/3} \times 2^{0.25(scfb+offset-16)-25} \quad (3)$$

(2)와 (3)에서 부호는 무작위 부호 발생기^[1]에 의해 결정되어 MDCT 계수가 잡음 신호로 복원된다. 위와 같이 잡음 채움 기법을 적용하기 위해, 부호기에서는 3비트의 잡음 레벨과 5비트의 잡음 오프셋을 추정하고, 비트열의 fd_channel_stream 구조체에 이 두 값을 포함시켜 전송한다.

III. Intelligent gap filling

IGF^[4]는 MDCT 영역에서 고주파 대역을 확장하는 기술로 MPEG-H 3DA^[1]에 처음으로 도입된 부호화 도구이다. SBR이 QMF 영역에서 대역 확장하는 것이 비해 IGF는 낮은 복잡도와 짧은 지연 시간을 요구하지만 동등한 수준의 대역 확장 성능을 보인다는 장점이 있다.

IGF는 MDCT 영역에서 IGF 시작 주파수와 IGF 정지 주파수를 설정한다. IGF 시작 주파수까지의 저주파 영역, 즉 핵심 대역은 기존의 MDCT 기반 파형 부호화 기법을 이용하는 핵심 부호기 (core codec)로 부호화한다. 반면에 IGF 시작 주파수부터 정지 주파수까지의 IGF 대역은 IGF 기법, 즉 복원된 핵심 대역을 복사하되, 적은 비트량으로 표현되는 대역 확장 파라미터들을 이용하여 복원한다.

보다 구체적으로, IGF에서는 먼저 IGF 대역을 SFB 단위로 분할하되 4개 이하의 패치들로 분할한다. 각각의 IGF

대역 패치에 대응되는 입력 패치는 IGF 시작 주파수 이하의 복원된 핵심 대역에서 4개의 후보 영역을 설정하고 그 후보 영역 중에서 최적의 영역으로 결정한다. 입력 패치를 구성하기 위해 복원된 핵심 대역 신호는 IGF 대역 패치별로 별도로 복사된다. 만약 독립 잡음 채움 기능이 설정되었다면 복사된 신호마다 개별적으로 잡음 채움을 수행한다. 그 후 IGF 백색화 (whitening)와 시간 영역 잡음 평활화 (temporal noise flattening) 등을 선택적으로 수행하여 입력 패치를 얻는다. 이와 같이 처리된 입력 패치에 있는 신호의 크기를 원래의 IGF 대역 신호의 크기와 맞추기 위해, 3DA에서는 IGF 대역에서 1개 혹은 2개 SFB 단위로 IGF 레벨을 전송한다. 이 IGF 레벨을 참조하여 입력 패치는 그 크기가 조정된 다음 지정된 IGF 대역 패치로 복사된다.

IV. 제안한 잡음 채움 기법

일반적으로 FD 모드 3DA 부호기에서는 IGF 시작 주파수까지의 저주파 영역을 핵심 부호기로 부호화하고, 그 이상의 고주파 영역을 IGF가 담당한다. 핵심 부호기가 부호화하는 대역의 끝을 최대 SFB, 즉 MAX_SFB라는 비트열 요소로 표현하여 전송하는데, MAX_SFB는 IGF 시작 주파수에 대응시켜 부호화한다. 이때 IGF 시작 주파수 아래에 0으로 양자화되는 주파수 성분들이 있다면 식 (2) 혹은 (3)으로 잡음 채움이 수행되어 복원될 것이다. 이러한 경우 삽입된 잡음은 귀에 들릴 정도의 큰 신호로 복원될 수 있어 주관적 품질에 큰 영향을 줄 수 있다. 이를 해결하기 위해 핵심 대역은 IGF 시작 주파수 이하로 제한하되, 부호화 마지막 단계에서 MAX_SFB를 재설정하여 핵심 대역의 끝 부분에 모두 0으로 양자화되는 SFB 들은 부호화 영역에서 삭제한다. 이러한 MAX_SFB 재설정은 핵심 대역 끝 부분에 잡음 채움을 수행하는 것을 막아, 과도한 잡음 채움으로 인한 주관적 음질 저하를 막을 수 있다.

다음으로, 각 IGF 대역 패치를 위한 입력 패치는 개별적으로 복원된 핵심 대역을 복사하여 구성한다. 이때, 복원된 핵심 대역이 3.5KHz 이하이거나 위에서 설정한 MAX_SFB 이상이라면 0으로 양자화된 값은 잡음 채움이 수행되는 영역 밖이어서 그 값이 0으로 유지되어 복원된

다. 입력 패치에 발생하는 이 스펙트럼 공백은 이에 대응되는 IGF 대역에도 그대로 남아있게 된다. 이는 IGF 대역 내 신호의 크기가 IGF 레벨에 관계없이 스펙트럼 공백으로 남아있을 수 있다는 것을 의미한다. 이러한 단점을 해결하기 위해 본 논문에서는 개별적인 입력 패치를 구성할 때 모든 주파수 영역에 대해 잡음 채움 기법을 적용한다. 이 과정을 통해 IGF 대역에서 스펙트럼 공백이 발생하지 않아 주관적 음질을 개선할 수 있도록 하였다. 이 과정은 MAX_SFB 재설정과 달리 부호기의 수정이 필요하여 표준의 수정이 필요하다.

V. 모의 실험

본 절에서는 제IV절에서 설명한 제안한 잡음 채움 기법의 성능을 기존의 3DA의 성능과 비교한다. 실험에 사용되는 오디오 코덱은 3DA 부호화기의 FD 처리부의 구문에 따라 부호화된다. FD 처리부에서는 TNS (temporal noise shaphing), M/S (mid/side), IGF, 잡음 채움 등의 부호화 도구를 포함한다. 실험에서는 [5]의 SW에 IGF와 잡음 채움 도구를 추가한 3DA 코덱 SW를 사용하여 24Kbps의 전송률로 부호화를 수행하였다. IGF는 시작 주파수를 5.25KHz, 끝 주파수를 15KHz로 설정하였다. 실험 샘플로는 3분 길이의 all_stereo.wav를 사용하였다. 이 샘플은 스테레오 신호를 48KHz로 표본화한 es01, salvation, louis_raquin_15, te15, wedding_speech, music1, tel_mg54_speech, music3, twinkle_ff51, speechOverMusic_1, speechOverMusic_4, HarryPotter, arirang_speech, lion 등 14개 파일을 순서대로 연결한 것으로 부호화 성능을 적절히 평가할 수 있을 정도로 다양한 특성을 보유하면서 충분한 길이를 가지고 있다.

부호화 성능을 비교하기 위해 5명의 오디오 전공자를 대상으로 MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) 테스트^[6]를 시행하였다. MUSHRA 테스트에서는 기존의 잡음 채움 및 IGF를 사용하는 방법 (3DA로 표기)과 수정된 잡음 채움 및 IGF를 사용하는 방법 (proposed로 표기)과 함께 원본 오디오 ref와 3.5KHz로 대역 제한된 앵커 신호 lp35 등 총 4개의 오디오 신호들을 비교하였다. 테스트에서는 all_stereo.wav를 부호화한 다음 성능 비교가 비교적 용이한 음성 샘플인 arirang, salvation,

wedding_speech 등 3 샘플에 해당하는 부분을 추출하여 평가하였다. 그런 다음 개별적인 3 샘플과 이들을 통합한 자료에 대한 음질 평가 결과의 대푯값 및 95% 신뢰구간을 그림 1에 제시하였다. 그림 1을 참조하면, 3DA 및 제안한 방법은 24Kbps의 낮은 전송률에서도 앵커 신호인 lp35 보다 주관적 품질이 높은 것으로 나타났다. 특히 제안한 방법은 3DA에 비해 arirang, salvation, wedding_speech에 대해 대푯값이 각각 약 9점, 8점, 12점 높게 나타났다. 3 샘플 결과를 통합한 그래프의 마지막 부분을 참조하면 제안한 방법은 3DA에 비해 약 10점 정도 높은 결과를 보였고, 95% 신뢰구간도 겹치지 않음을 알 수 있다. 따라서 제안한 방법은 기존의 방법보다 부호화 성능을 개선함을 알 수 있었다.

그림 2에서는 3DA와 제안한 기법으로 부호화하였을 때 복원된 all_stereo.wav의 왼쪽 스펙트럼을 도시하였다. 그림에서 IGF 시작 주파수를 오른쪽 화살표로 표기하였다. 그림 2를 참조하면, 3DA에서는 핵심 대역뿐만 아니라 IGF 대역에서도 스펙트럼 공백이 많이 나타남을 알 수 있다. 반면에 제안한 기법에서는 IGF 대역이 스펙트럼 공백 없이 매우 충실히 복원됨을 알 수 있다. 그리고 IGF 시작 주파수 이하에서 스펙트럼 공백이 잡음 채움 없이 나타나 지나친 잡음 채움으로 인한 부자연스러움을 회피함을 보여준다.

VI. 결론

본 논문에서는 IGF를 채용한 3DA 부호기의 FD 처리부에서 새로운 잡음 채움 기법을 제안하였다. 제안한 기법의 성능은 MUSHRA 테스트로 그 성능을 분석한 결과, 기존의

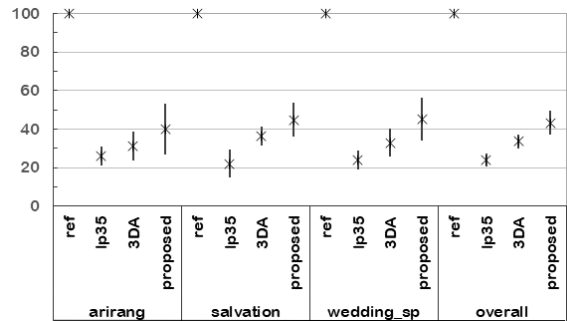


그림 1. MUSHRA 테스트 결과
Fig. 1. Results of the MUSHRA test

방법에 비해 높은 주관적 품질을 보임을 알 수 있었다. 본 논문에서의 MUSHRA 테스트에서는 제안한 기법이 60점 이하의 낮은 주관적 품질을 보였다. 추가적인 연구를 수행한다면, 24Kbps 이하의 저 전송률에서도 MPEG-H 3DA 부호기의 성능을 좀 더 고도화시킬 수 있을 것으로 기대한다.

참고 문헌 (Reference)

- [1] ISO/IEC FDIS 23008-3:2019(E), *Information Technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 3: 3D Audio*, 2019.
- [2] N. Rettelbach, et al., "Audio Encoder, Audio Decoder, Methods for Encoding and Decoding an Audio Signal, and a Computer Program," US Patent 9449606 B2, Fraunhofer-Gesellschaft, 2016.
- [3] S. Disch, et al., "Noise Filling in Perceptual Transform Audio Coding," WO Patent 2014/118176 A1, Fraunhofer-Gesellschaft, 2014.
- [4] S. Disch, et al., "Intelligent Gap Filling in Perceptual Transform Coding of Audio," *141th AES Convention*, LA, USA, Paper 9661, Sept. 2016.
- [5] Rin Chul Kim "A 3D Audio Core-Codec Employing an Improved Buffer Control Method" *JBE* Vol. 25, No. 2, pp 233-241, March 2020.
- [6] ITU-R BS.1534-1, *Method for the Subjective Assessment of Intermediate Sound Quality(MUSHRA)*, 2011.

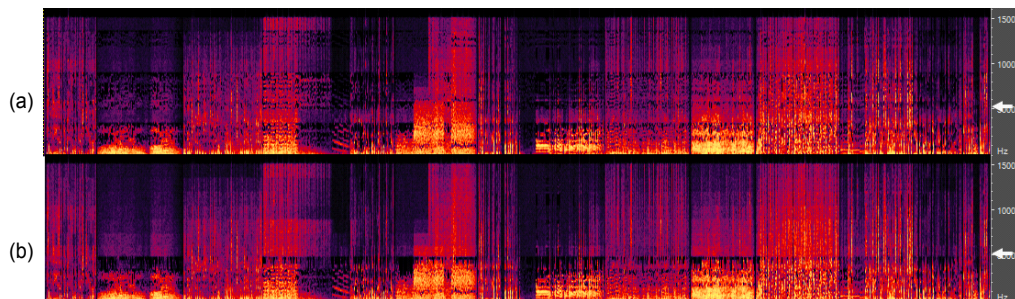


그림 2. (a) 3DA, (b) 제안한 방법으로 all_stereo.wav을 부호화하여 복원한 왼쪽 채널 스펙트럼
Fig. 2. Left channel spectra of the all_stereo.wav reconstructed by (a) 3DA, (b) the proposed method