

일반논문 (Regular Paper)

방송공학회논문지 제25권 제2호, 2020년 3월 (JBE Vol. 25, No. 2, March 2020)

<https://doi.org/10.5909/JBE.2020.25.2.233>

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

향상된 버퍼 제어 방법을 사용한 3D 오디오 핵심 부호화기

김인철^{a)†}

A 3D Audio Core-Codec Employing an Improved Buffer Control Method

Rin Chul Kim^{a)†}

요 약

본 논문에서는 3DA 핵심 부호화기의 FD 처리부의 성능 향상을 위해 새로운 버퍼 제어 기법을 제안하였다. 제안한 버퍼 제어 기법은 3GPP의 심리음향 모델과 스펙트럼 공백 회피 기법을 포함한 울-왜곡 제어부를 채택하였다. 그런 다음, 활동성이 높은 프레임에 충실한 비트 할당이 이루어질 수 있도록 3GPP 버퍼 제어 방법을 수정하였다. 본 논문에서 제안한 버퍼 제어 방법의 성능은 MUSHRA 테스트를 통해 평가되었고, 그 결과 3DA RM9 및 3GPP AAC의 버퍼 제어 방법보다 우수한 성능을 보임을 입증한다.

Abstract

In this paper, a new buffer control method is proposed for improving the performance of the frequency domain part of the 3D audio (3DA) core codec. For the proposed buffer control method, we first combine the 3DA RM9 with the 3GPP AAC buffer control method which includes the psychoacoustic model and rate-distortion control process with the spectral hole avoidance algorithm. Then, we revise the 3GPP buffer control method so as to achieve a faithful bit allocation to the frames with higher activity. With the MUSHRA test, we prove that the proposed buffer control method demonstrates better performance than the 3DA RM9 and 3GPP AAC.

Keyword : audio coding, 3DA RM, core codec, 3GPP AAC, buffer control method

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 2017 Research Fund of the University of Seoul.

• Manuscript received December 20, 2019; Revised January 31, 2020; Accepted January 31, 2020.

1. 서 론

오디오 신호는 방대한 데이터 양으로 인해 효과적인 저장 및 전송을 위한 오디오 압축 즉 오디오 부호화에 대한 연구가 수십년전부터 이루어졌다. 오디오 부호화는 MPEG-2 AAC (advanced audio coding)^[1]부터 MDCT에 기반한 오디오 부호화로 체계적인 구조를 가지게 되었다. 이러한

MDCT 기반 오디오 부호화는 MPEG에서 MPEG-4 오디오^[2], USAC (unified speech and audio coding)^[3], 최근의 3DA (3D audio)^[4]를 통해 그 기능이 확장되고, 다양한 응용 분야에 적용될 수 있도록 발전했다.

USAC 및 3DA 등에서 지금까지 여러 가지 부호화 도구들이 추가되었지만, MDCT 기반 오디오 부호화에서 가장 기본적인 부호화 도구는 심리음향 모델, MDCT, TNS (temporal noise shaping), M/S (mid/side) 부호화, 그리고 양자화 및 무손실 부호화 등을 포함하는 울-왜곡 제어처리부 (rate-distortion control process) 등으로 볼 수 있다. 다른 부호화 도구들은 이 기본적인 부호화 도구들의 기능을 확장하거나, 그들을 기반으로 하여 향상된 성능을 도모하기 위한 것으로 볼 수 있다. 따라서 기본적인 부호화 도구의 성능이 전체 MDCT 기반 오디오 부호화의 성능을 좌우하게 된다. 특히, 울-왜곡 제어는 심리음향 모델과 결합하여 입력 오디오 신호의 활동성과 버퍼 공백도 (buffer vacancy)에 따라 동적으로 비트량을 할당하는 부분으로, 부호화 성능에 가장 크게 기여하는 부분 중의 하나이다. 본 논문에서는 심리음향 모델과 결합된 울-왜곡 제어를 통해 비트 할당하는 방법을 버퍼 제어 방법 (buffer control method)라고 부르며, MDCT 기반 오디오 부호화의 성능을 향상시킬 수 있는 새로운 버퍼 제어 방법을 제안하고자 한다.

본 논문에서는 MDCT 기반 오디오 부호화의 최신 표준인 3DA^[4] 구문을 따르는 3DA RM9 (reference model 9)을 기준으로 부호화한다. MDCT 기반 부호화기 중에서 3GPP AAC^[5]는 부호화기의 구조가 공개되어 있고 현재 가장 널리 사용되는 부호화 중의 하나일 뿐만 아니라 우수한 성능을 보이는 것으로 알려져 있다^[7]. 본 논문에서는 먼저 3GPP AAC의 버퍼 제어 기법을 3DA RM9에 적용한다. 그리고 신호의 활동성에 따라 적응적으로 비트 할당을 수행하는 버퍼 제어 방법을 제안한다. 제안한 버퍼 제어 방법은 모의 실험을 통해 기존의 RM9 및 3GPP AAC의 버퍼 제어 방법과 그 성능을 비교한다.

본 논문의 구성은 다음과 같다. 2절에서는 MDCT 기반 오디오 부호화기를 간략히 살펴보고, RM9과 3GPP 등 기존의 버퍼 제어 기법들을 설명한다. 3절에서는 본 논문에서 제안하는 버퍼 제어 기법을 소개하고, 4절에서는 모의 실험을 통해 제안하는 기법의 성능을 비교 평가한다. 마지막으

로 5절에서는 본 논문의 결론을 제시한다.

II. 기존의 버퍼 제어 방법

1. FD 처리부 개관

MDCT 기반 오디오 부호화 기법은 MPEG-2 AAC^[1] 이래로 MPEG-4 오디오^[2], USAC^[3] 등에 적용되었다. 최근에 표준으로 완성된 MPEG 3DA^[4]에서도 핵심 부호화기 FD (frequency domain) 처리부로 MDCT 기반 오디오 부호화 기법이 포함되었다. MDCT 기반 오디오 부호화 기법의 기본적인 구조는 그림 1과 같이 볼 수 있다. 그림 1에 제시한 바와 같이 MDCT 기반 오디오 부호화 기법의 기본적인 부호화 도구는 심리음향 모델, 필터 बैं크, TNS, M/S, 그리고 울-왜곡 제어 처리부 등이다. 울-왜곡 제어 처리부는 scale factor로 신호의 크기를 조절하는 scaling, 잡음 채움 (noise filling)을 포함하는 양자화, 그리고 산술 부호화로 대표되는 무손실 부호화 등으로 구성된다.

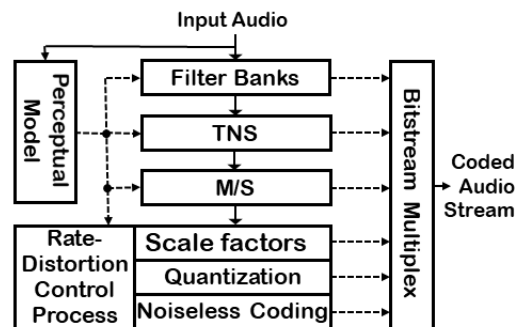


그림 1. 기본적인 MDCT 기반 오디오 부호화기의 블록선도
Fig. 1. Block diagram of a basic MDCT based audio coder

MDCT 기반 오디오 부호화 기법은 AAC, USAC, 3DA 등의 비트열 구문에 따라 구현될 수 있는데, 본 논문에서는 가장 최근의 표준인 3DA 구문에 따라 부호화한다. 3DA 구문에 따르면, 입력 오디오는 오디오 채널 수에 따라 미리 정해진 방법으로 하나 혹은 2개의 채널 단위로 SCE (single channel element), CPE (channel pair element), 혹은 LFE (low frequency effect element) 등으로 구분한다. 그런 다

음, 각 SCE, CPE, 혹은 LFE 들에 대해 개별적으로 3DA 핵심 부호화기를 실행시켜 부호화된 비트열을 얻는다. 최종적으로 각 비트열을 미리 정해진 순서대로 결합하여 통합된 하나의 비트열을 산출한다.

그림 1에 제시한 부호화 도구 이외에 USAC과 3DA는 추가적인 부호화 도구들을 도입하였다. 예를 들어 USAC^[3]은 time-warped MDCT, 복소 스테레오 예측 (complex stereo prediction), 잡음 채움 등을 도입하였고, 3DA^[4]는 IGF (intelligent gap filling), MCT (multi-channel coding tool), FDP (frequency domain prediction) 등의 부호화 도구들을 추가하였다. 그러나 USAC과 3DA에서 추가된 부호화 도구들은 그림 1에 제시한 기본적인 부호화 도구들의 기능을 확장하거나, 그들을 기반으로 하여 향상된 성능을 제공하기 위한 것이다. 따라서 그림 1에 제시된 기본적인 부호화 도구의 성능이 전체 MDCT 기반 오디오 부호화의 성능을 좌우하게 된다. 특히, 심리음향 모델과 결합된 울-왜곡 제어 즉 버퍼 제어 방법은 부호화 성능에 큰 영향을 준다. 다음 절에서는 그림 1의 기본적인 부호화 도구들로 구성된 오디오 부호화기에서 버퍼 제어 방법에 관해서 좀 더 구체적으로 살펴본다.

2. 버퍼 제어

3DA 표준의 FD 처리부에서 변환 계수들은 양자화된 값과 scale factor로 표현되며 이들은 효율적인 정보 표현을 위해 산술부호화 혹은 Huffman 부호화 등 무손실 부호화된다. 무손실 부호화된 변환 계수들은 다른 부가 정보들과 함께 결합되어 프레임 단위로 비트열 형태로 표현되어 전송 혹은 저장된다.

3DA 표준^[4]을 이용하여 부호화할 때 프레임당 산출되는 비트열의 길이, 즉 프레임당 비트량은 해당 프레임의 특성에 따라 가변적이다. 예를 들어 무음 구간에서는 프레임당 비트량은 0비트에 가까울 것이고, 오케스트라의 클라이맥스 구간처럼 다양한 음원이 큰 크기로 섞여 있는 경우에는 프레임당 비트량이 매우 많을 것이다. 이러한 상황에서 오디오 신호를 고정된 비트율로 효과적으로 표현하려면 버퍼 혹은 비트 저장소를 설치하는 것이 필요하다. 프레임당 산출되는 비트들은 먼저 버퍼에 저장하고, 버퍼에서 일정한

속도로 비트들을 출력함으로써 고정된 비트율을 얻을 수 있다. 버퍼를 이용한 데이터 율 조정, 즉 버퍼 제어에서는 버퍼의 크기와 버퍼 제어 방법이 부호화 효율에 큰 영향을 준다.

3DA 표준에서 버퍼의 크기 $max_bitreservoir_bits$ 는 유효채널 수에 6144비트를 곱한 값으로 정의한다. 예를 들어 2채널 스테레오 신호의 경우 유효 채널 수가 2이므로 버퍼의 크기는 12288비트가 된다. 만약 신호의 표본화 주파수가 48KHz이고 프레임 길이가 1024라고 할 때 표준에 의한 버퍼 크기에서 주어질 수 있는 최대 비트율은 프레임당 6144비트이고 이는 채널 당 288Kbits/sec에 해당한다. 실제 해당 프레임의 부호화를 통해 bit_gen 만큼의 비트가 발생하였다면 버퍼에 비어있는 비트량 $bitresBits$ 는 다음과 같이 갱신된다.

$$bitresBits += avgBits - byte_aligned(bit_gen) \quad (1)$$

여기서 $byte_aligned()$ 함수는 필요한 경우 비트 추가를 통해 byte 단위로 비트열 길이를 맞추는 함수이다. $avgBits$ 는 프레임당 평균 비트량으로, 프레임의 길이가 1024이고 표본화 주파수가 48KHz이고, 전송률이 42Kbps일 때 프레임 당 평균 비트량은 896비트이다.

2.1 RM9의 버퍼 제어

3DA RM9^[6]에서는 매우 간단한 버퍼 제어 방법을 채택하였다. 그림 2에 제시한 바와 같이 채널 당 6144비트 크기

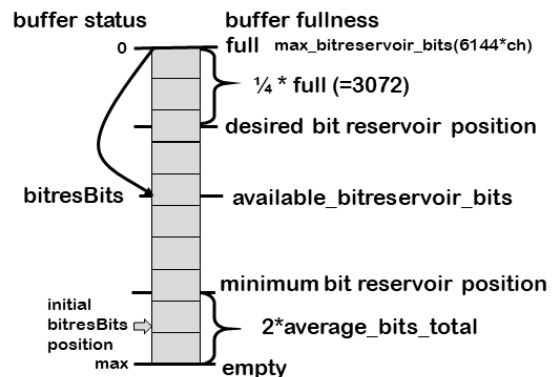


그림 2. MPEG 참조 모델 9에 설치된 버퍼
 Fig. 2. The buffer installed at the MPEG reference model 9

의 버퍼에서 버퍼 크기의 1/4 되는 위치를 목표 비트 저장소 위치로 설정한다. 현재 버퍼에 비어있는 비트량인 가용 비트 저장소 비트량을 $bitresBits$ 라고 할 때, 현재 프레임에 할당되는 비트량 $num_bits_available$ 은 다음 식과 같이 주어진다.

$$num_bits_available = avgBits + bitresBits - 0.25 \cdot max_bitreservoir_bits \quad (2)$$

위 식을 통해 매 프레임에 비트를 할당하고, 오디오 부호기가 성공적으로 할당된 비트량만큼 비트열을 산출한다면, 버퍼는 목표 비트 저장소 위치만큼 채워질 것이다.

RM9에서는 프레임에서 산출되는 비트량을 할당된 비트량으로 제한하기 위해 2 루프 반복 알고리즘을 이용해 SFB (scale factor band) 별로 scale factor를 조정한다. 안쪽 루프에서는 부호화 후 발생하는 왜곡과 문턱값 (threshold)의 비를 일정하게 만들기 위해 그 비가 큰 SFB의 scale factor를 증가시킨다. 여기서 문턱값은 매우 간단한 심리음향 모델을 통해 얻는데, SFB의 에너지를 14dB만큼 감소시킨 값으로 설정된다. 바깥쪽 루프에서는 공통 (common) scale factor를 증가시키고 안쪽 루프를 호출하는 과정을 반복하면서 부호기에서 산출되는 비트량이 할당된 비트량 이하가 되도록 조정한다.

만약 실제 산출되는 비트량이 프레임에 할당된 비트량에 지속적으로 미달된다면 버퍼 공백 상태 (buffer empty state)에 도달할 수 있다. RM9에서는 버퍼 공백 상태를 방지하기 위해 프레임당 평균 비트량 $avgBits$ 의 2배 크기의 최소 비트 저장소 위치를 설정한다. 만약 버퍼 충만도가 최소 비트 저장소 위치 아래로 떨어지면 강제 비트 삽입 (bit stuffing)을 수행하여 버퍼 충만도가 이 위치 이상이 되도록 만든다.

2.2 3GPP AAC의 버퍼 제어

3GPP AAC^[5]에서는 버퍼 상태와 프레임의 복잡도를 동시에 고려하여 현재 프레임에 할당되는 비트량을 결정하는 매우 정교한 버퍼 제어 방법을 도입하였다. 먼저 버퍼 상태로부터 비트 할당의 하한선을 나타내는 bit_save 와 비트 할

당의 상한선을 나타내는 bit_spend 를 산출한다. 그리고 각 프레임의 활동성은 PE (perceptual entropy)로 추정한다. 이전 프레임까지의 PE들을 참조하여 PE의 범위를 최소값 $peMin$ 과 최대값 $peMax$ 로 추정한다. 그런 다음, bit_save 와 bit_spend 를 각각 $peMin$ 과 $peMax$ 에 대응시킨다. $peMin$ 과 $peMax$ 와 현재 프레임의 PE 간의 거리에 비례하여 f 값을 계산한다. 이 f 값과 프레임당 평균 비트량 $avgBits$ 을 참조하여 현재 프레임의 목표 비트량 $desiredBits$ 를 얻는다.

3GPP AAC의 버퍼 제어 방법을 좀더 구체적으로 설명하면 다음과 같다. 주어진 프레임의 활동성은 SFB 단위로 에너지, 문턱값, 그리고 유효 계수 수 (number of effective lines) 등을 참조하여 추정된다. 주어진 SFB sfb 에서의 에너지는 다음과 같이 MDCT 계수 $X[j]$ 의 제곱 합으로 계산한다.

표 1. 3GPP AAC에서 pe 의 최대값과 최소값 갱신 방법
Table 1. Methods for updating the maximal and minimal values of pe in the 3GPP AAC

3GPP	$pe < peMin$	$peMin < pe < peMax$	$pe > peMax$
$peMin +=$	$0.14 \times (pe - peMin)$	$0.3 \times (pe - peMin)$	$0.3 \times (pe - peMax)$
$peMax +=$	$0.07 \times (pe - peMin)$	$0.07 \times (pe - peMax)$	$1 \times (pe - peMax)$

$$en[sfb] = \sum_{j \in sfb} (X[j])^2 \quad (3)$$

문턱값 $thr[sfb]$ 은 에너지를 -29dB만큼 감축시킨 다음, 주변 SFB로 확산(spreading)하고, 직전 프레임 문턱값을 참조하여 pre-echo 제어를 수행하여 얻는다. 유효 계수 수는 다음 식과 같이 얻는다.

$$NLines[sfb] = \frac{\sum_{j \in sfb} |X[j]|^{1/2}}{\left(\frac{1}{\sum_{j \in sfb} 1} \sum_{j \in sfb} (X[j])^2 \right)^{1/4}} \quad (4)$$

PE는 SFB 별로 산출되는 에너지, 문턱값, 유효 계수 수를 식(5)와 같이 조합하여 계산한다. 여기서 $offset$ 은 채널당

$$pe = offset + \sum_{sfb \in (en[sfb] > thr[sfb])} NLines[sfb] \times \left(C_2 + C_3 \frac{en[sfb]}{thr[sfb]} \right) \quad (5)$$

비트율이 16Kbps 이하는 50, 32Kbps 이상은 0이며 그 사이의 비트율에서는 두 점을 선형 보간한 값으로 정해진다. C_2 는 $en[sfb]$ 가 $thr[sfb]$ 보다 8배 이상 크면 0이고 그렇지 않으면 $\log_2(2.5)$ 이다. C_3 은 $en[sfb]$ 가 $thr[sfb]$ 보다 8배 이상 크면 1이고 그렇지 않으면 $1 - \log_2 2.5 / \log_2 8$ 이다.

매 프레임의 부호화가 끝나면, pe 의 범위를 표 1과 같이 $peMin$ 과 $peMax$ 값으로 추정한다. 이때 pe 의 범위 $peMax - peMin$ 은 pe 값의 1/6 이하가 되도록 강제한다. 프레임당 목표 비트량은 프레임당 평균 비트량 $avgBits$ 에 $1 + f$ 를 곱하여 얻는데, f 는 그림 3과 같이 가용 비트 저장소 비트량 $bitresBits$ 와 주어진 프레임의 pe 의 함수로 주어진다. 그림 3을 살펴보면, 가용 비트 저장소 비트량의 범위 $[0.95, 0.2]$ 는 bit_save 와 bit_spend 에서 각각 $[0.05, -0.3]$ 과 $[0.4, 0.1]$ 범위의 값들로 대응된다. 실제 bit_save 와 bit_spend 값은 각 범위를 이어주는 직선 상에서 현재의 $bitresBits$ 에 대응하는 값들로 산출되며 f 값은 이 bit_save 와 bit_spend 값을 연결하는 직선상에 위치하게 된다. 마지막으로 bit_save 값에 $peMin$ 값을 대응시키고, bit_spend 값에 $peMax$ 값을 대응시켰을 때 두 점을 연결하는 직선상에서 현재 프레임의 pe 에 대응되는 값을 f 값으로 추정한다.

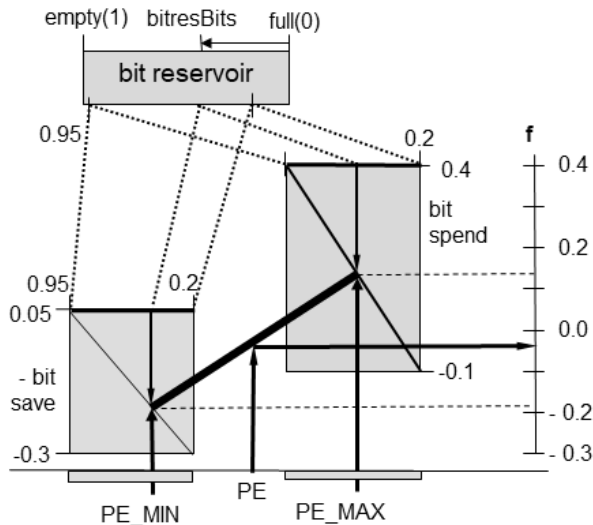


그림 3. 3GPP AAC에서 채택한 버퍼 제어 방법
Fig. 3. The buffer control method in the 3GPP AAC

3GPP AAC에서는 여기서 구한 f 를 참조하여 목표 PE 값을 다음과 같이 설정한다.

$$PE_r = \alpha_c \times (1 + f) \times 1.18 \times avgBits \quad (6)$$

주어진 프레임의 pe 가 목표값 PE_r 이 될 때까지 각 SFB에서 발생하는 잡음의 loudness가 균일해질 수 있는 방향으로 문턱값 $thr[sfb]$ 를 증가시킨다. 이때 스펙트럼 공백 회피 (spectral hole avoidance) 기법을 적용하여 복원된 오디오 신호에 과도한 스펙트럼 공백이 발생하는 것을 방지한다. 이와 같은 과정을 통해 문턱값이 결정되면 이를 참조하여 scale factor를 다음과 같이 추정한다.

$$sf[sfb] = \text{floor} \left(8.8585 \times \log_{10} \left(6.75 \times thr[sfb] / \sum_{k \in sfb} \sqrt{|X[k]|} \right) \right) \quad (7)$$

한편, 가용 비트 저장소 비트량 $bitresBits$ 가 주어졌을 때 현재 프레임 부호화에 사용할 수 있는 최대 비트량 $num_bits_available$ 은 다음과 같이 주어진다.

$$num_bits_available = avgBits + bitresBits - 7 \quad (8)$$

만약 위에서 구한 $sf[sfb]$ 로 부호화한 비트량이 $num_bits_available$ 보다 적다면 최종적으로 그 $sf[sfb]$ 로 부호화한다. 그렇지 않다면 실제 부호화한 비트량이 $num_bits_available$ 보다 적어질 때까지 모든 $sf[sfb]$ 들을 1씩 증가시키면서 부호화를 반복한다. 좀 더 자세한 3GPP AAC의 버퍼 제어 기법은 3GPP AAC 규정^[5]을 참조하길 바란다. 3GPP AAC의 버퍼 제어 방법은 MPEG 참조 모델의 버퍼 제어 방법에 비해 매우 복잡하지만, 그 성능이 우수하다.

III. 제안한 버퍼 제어 방법

3GPP AAC의 버퍼 제어 방법^[5]은 3DA RM의 버퍼 제어 방법^[6]에 비해 우수한 성능을 보인다. 그러나 3GPP 버퍼

표 2. 제안한 방법에서 pe 의 최대값과 최소값 갱신 방법

Table 2. Methods for updating the maximal and minimal values of pe in the proposed method

proposed	$pe < peMin$	$peMin < pe < peMax$	$pe > peMax$
$peMin +=$	$0.03 \times (pe - peMin)$	$0.05 \times (pe - peMin)$	$0.05 \times (pe - peMax)$
$peMax +=$	$0.015 \times (pe - peMin)$	$0.015 \times (pe - peMax)$	$0.1 \times (pe - peMax)$

제어 방법을 3DA 핵심 부호기의 FD 처리부에 적용하였을 때, 버퍼의 상태가 충만 상태에 근접하고, 신호의 활동성 범위를 결정하는 파라미터들을 급격히 변화시키기 때문에 신호의 활동성에 충분히 대응하지 못한다는 단점이 있다. 따라서 본 논문에서는 다음과 같이 3GPP의 버퍼 제어 방법을 수정한다.

첫 번째로, 평균적인 버퍼의 상태를 충만 상태에서 중간 상태로 옮긴다. 3GPP 버퍼 제어 방법은 버퍼의 상태를 충만에 근접한 상태에 머물게 한다. 이러한 경우 가용 비트량이 적어지므로, 높은 활동성을 가지는 신호에 대해 필요한 만큼의 충분한 비트량을 제공하지 못할 수 있다는 단점이 발생한다. 본 논문에서는 버퍼의 상태를 중간 상태로 유지하기 위해 두 가지 방법을 사용한다. 먼저 3GPP 방법에서는 식 (6)과 같이 목표 PE 값을 설정하였는데, 이 경우 초기의 목표 PE 값들이 커져서 신호의 활동성에 비해 많은 비트량을 할당하게 되고 이에 따라 버퍼가 충만 상태에 근접하게 된다. 부호화가 진행됨에 따라 f 값은 약 -0.3 정도의 매우 낮은 값으로 변경되어 목표 PE 값은 평균 비트량에 근접하지만, 낮은 f 값으로 인해 목표 비트량은 매우 적어진다. 본 논문에서는 이러한 단점을 완화하기 위해 목표 PE 값을 다음 식과 같이 3GPP 방법에서 정한 것보다 조금 낮은 값으로 설정한다.

$$PE'_r = \alpha_c \times (1 + f) \times 0.82 \times avgBits \quad (9)$$

다음으로 현재 프레임에 할당할 수 있는 최대 비트량을 재설정한다. 3GPP 기법에서는 식 (8)과 같이 최대 비트량을 가용 비트량으로 설정하였다. 이 경우 현재 프레임이 버퍼에 남아있는 모든 비트들을 소비할 수 있으므로, 다음에 따라오는 프레임이 높은 활동성을 보인다면 그 프레임에 충분한 비트를 할당할 수 없을 수도 있다는 단점이 발생한

다. 따라서 본 논문에서는 최대 비트량을 다음 식과 같이 목표 비트량으로 설정한다.

$$desiredBits = (1 + f) \times avgBits \quad (10)$$

즉, 식 (9)와 같이 목표 비트량의 약 0.82배만큼 목표 PE 값을 설정하여 문턱값 등 부호화 파라미터를 설정하고, 실제 양자화 및 무손실 부호화하여 산출된 비트량이 목표 비트량 이하인지를 점검한다. 이 과정을 통해 버퍼는 충만 혹은 공백 상태에 치우치지 않고 중간 상태를 유지함으로써 높은 활동성을 보이는 프레임에 대해 항상 필요한 만큼의 비트량을 제공할 수 있게 된다.

두 번째로 비트량 제어를 3GPP 방법과 같이 개별적인 채널 단위로 수행하지 않고 프레임 내 모든 채널들에 대해 동시에 수행한다. 3GPP 방법에서는 현재 프레임에 가용한 최대 비트량을 각 채널의 PE의 비율로 분할하여 채널별 가용 비트량을 계산한다. 그런 다음 개별적인 채널 별로 산출되는 비트량을 채널별 가용 비트량 내로 제한한다. 이 방법은 PE가 채널별 비트량과 정확히 비례 관계에 있지 않기 때문에 채널 간에 적절한 비트 할당이 수행되지 않을 수 있다는 단점이 있다. 본 논문에서는 프레임 내의 모든 채널에 대해 비트량 제어를 동시에 수행한다. 즉, 현재 프레임의 모든 채널에서 산출되는 비트량을 모두 합하여 현재 프레임에 대한 최대 비트량인 식 (10)의 목표 비트량과 비교한다. 만약 산출된 비트량이 목표 비트량보다 크면, 모든 채널의 scale factor를 1만큼 증가시켜 부호화하여 비트량을 다시 산출한다. 이 과정을 반복하여 모든 채널에서 산출된 비트량의 합이 목표 비트량보다 작게 만든다.

세 번째로, 활동성의 범위를 나타내는 $peMin$ 과 $peMax$ 의 갱신 속도를 수정하여 현재 프레임의 활동성을 보다 효과적으로 반영할 수 있도록 한다. 현재 프레임의 활동성은 과거의 프레임들의 pe 값들로부터 $peMin$ 과 $peMax$ 를 얻고 현재 프레임의 pe 값을 $peMin$ 과 $peMax$ 와 비교함으로써 추정한다. 3GPP에서 적용하는 표 1의 $peMin$ 과 $peMax$ 갱신 방법은 그 갱신 속도가 빨라 현재 프레임의 활동성을 효과적으로 표현하기가 어렵다. 따라서 표 2과 같이 갱신을 위한 승수들을 약 0.1~0.2배로 작게 하여 갱신 속도를 크게 낮추었다. 이를 통해 활동성이 높은 프레임이 반복해서 나

타날 때, 그 프레임들에게 지속적으로 많은 비트량을 할당할 수 있도록 한다.

네 번째로 그림 3에 제시한 비트 할당을 위한 파라미터들을 표 3과 같이 수정한다. 먼저, 버퍼 상태의 범위를 0.2~0.95에서 0.2~0.8로 조정하여 버퍼 충전 상태뿐만 아니라 버퍼 공백 상태로부터도 일정한 거리를 두도록 한다. 그리고 f 값의 하한값을 결정하는 bit_save 와 f 값의 상한값을 결정하는 bit_spend 범위를 넓힌다. 특히, 양의 범위를 크게 넓혀 활동성이 높아 PE 값이 커졌을 때 충분한 비트가 할당되도록 한다.

표 3. 비트 할당을 위한 버퍼 제어 파라미터의 수정
 Table 3. Revised buffer control parameters for bit allocation

	3GPP	proposed
range of buffer status	0.2 ~ 0.95	0.2 ~ 0.8
f values for bit save	-0.3 ~ 0.05	-0.3 ~ 0.4
f values for bit spend	-0.1 ~ 0.4	-0.1 ~ 1.2

IV. 모의 실험

본 절에서는 제3절에서 제시한 버퍼 제어 방법의 성능을 다른 2가지 방법과 비교한다. 모든 비교 대상 코덱들은 3DA 핵심 부호화기의 FD 처리부의 구분에 따라 부호화된다. 모든 코덱들은 부호화 효율을 위해 입력 스테레오 신호를 9KHz로 대역제한하였고, 전송률 40Kbps로 부호화한다. 부호화 도구 중에서 TNS와 M/S는 포함되고 IGF 등 다른 부호화 도구들은 포함하지 않는다. 첫 번째 비교 대상은 3DA RM9^[6]으로 II.2.1절에 설명한 버퍼 제어 방법을 사용한다. 두 번째 비교 대상은, 본 절에서 3GPP 방법으로 부르는데, 3DA RM9에 II.2.2절에서 설명한 3GPP^[5]의 버퍼 제어 방법을 적용한 것이다. 실험 샘플로는 3분 길이의 all_stereo.wav를 사용하였다. all_stereo.wav는 es01, salvation, louis_raquin_15, te15, wedding_speech, music1, tel_mg54_speech, music3, twinkle_ff51, speechOverMusic_1(SOM1), speechOverMuisic_4, HarryPotter, arirang_speech, lion 등 14개의 파일을 순서대로 연결한 것으로, 버퍼 제어 방법의 성능을 적절히 평가할 수 있을 정도로 다양한 특성을 보유

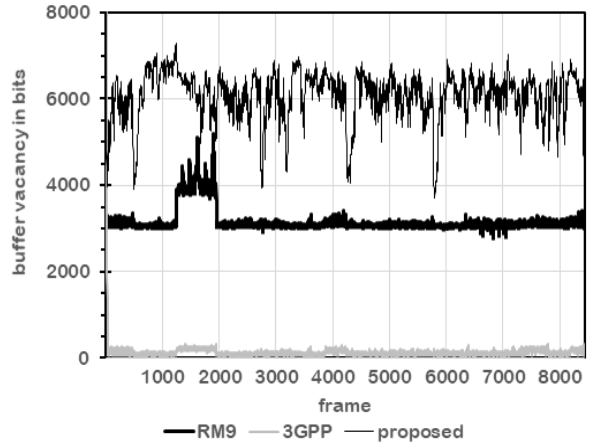


그림 4. 3가지 버퍼 제어 방법으로 부호화하였을 때의 버퍼 공백도
 Fig. 4. The buffer vacancies when coded by the 3 buffer control methods

하면서 충분한 길이를 가지고 있다.

그림 4에서는 3가지 버퍼 제어 기법으로 부호화하였을 때 버퍼 공백도를 보여준다. 실험 대상 오디오가 스테레오 신호이므로 버퍼 공백도의 최대치는 12288비트이고, 버퍼 공백도가 0이면 버퍼가 충전상태이다. RM9는 목표 비트 저장소 위치가 버퍼 크기의 1/4 즉 3072 비트이므로, 버퍼 공백도는 그림 4에서 3000비트 근방에 위치한다. 3GPP는 식 (8)과 같이 버퍼에 남아있는 모든 비트량을 현재 프레임에 할당할 수 있기 때문에 버퍼 공백도는 0 근방에 위치한다. 제안한 방법은 약 6000비트의 중간 상태로 버퍼 공백도를 유지하고 있음을 알 수 있다. 그리고 약 4000비트부터 7000비트까지 버퍼 공백도가 변화하는 것으로 보여 프레임의 활동성에 따라 각 프레임에 실제 할당되는 비트량이 달라짐을 알 수 있다. 한편, RM9, 3GPP, 제안한 방법으로 부호화하였을 때 실제 산출되는 비트량의 평균은 각각 40.00, 39.99, 40.01Kbps로, 3가지 방법 모두 충실하게 버퍼를 제어함을 알 수 있었다.

3가지 버퍼 제어 방법에 따른 부호화 성능을 비교하기 위해 8명의 오디오 전공자를 대상으로 MUSHRA 테스트^[8]를 시행하고 그 결과의 평균 및 95% 신뢰 구간을 그림 5에 제시하였다. MUSHRA 테스트에서는 RM9, 3GPP, proposed 등 3가지 버퍼 제어 방법으로 부호화한 결과와 함께, 원본 오디오 ref와 3.5KHz로 대역 제한된 앵커 신호 lp35

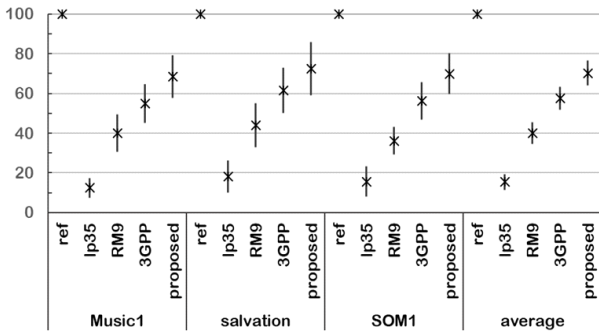


그림 5. MUSHRA 테스트 결과
Fig. 5. Results of the MUSHRA test

등 총 5개의 오디오 신호들을 비교하였다. 테스트에서는 all_stereo.wav를 부호화한 다음 Music1, salvation, 그리고 SOM1에 해당되는 부분을 추출하여 평가하였다. 그런 다음 각 신호에 대한 결과 및 그들의 평균을 그림 5에 제시하였다. 그림 5를 참조하면, RM9에 비해 3GPP 방법은 약 15~20 만큼 높은 결과를 보인다. 3GPP 방법은 RM9에 비해 향상된 심리 음향 모델과 스펙트럼 공백 회피 등으로 인해 주관적 품질이 크게 개선되는 것으로 보인다. 이에 더

하여 제안한 버퍼 제어 방법은 3GPP에 비해 약 10~15 정도 더 높은 결과를 보인다. 그림 5의 마지막에 보인 평균 결과를 보면, 제안한 방법은 RM9에 비해 약 30, 3GPP에 비해 약 13만큼 높은 점수를 보일 뿐만 아니라 신뢰구간이 겹치지 않음을 알 수 있다. 따라서 제안한 기법은 버퍼의 상태를 중간으로 유지하고 활동성이 높은 프레임에 충분한 비트량을 제공할 수 있도록 함으로써 주관적 품질이 개선됨을 알 수 있었다.

그림 6에서는 3가지 버퍼 제어 방법을 사용하여 부호화 하였을 때 복원된 all_stereo.wav의 왼쪽 스펙트럼을 도시 하였다. 본 실험에서는 9KHz 대역폭으로 부호화하므로 그림 6의 복원된 스펙트럼들은 최대 9KHz까지 스펙트럼이 펼쳐져 있다. 그러나 그림 6(a)에 보인 RM9의 결과는 약 3~5KHz로 대역 제한된 것처럼 보이는 부분들이 나타나며, 다수의 스펙트럼 공백이 보인다. 그림 6(b)의 3GPP 결과는 RM9에 비해 부호화되는 대역폭이 확장되었음을 알 수 있다. 그림 6(c)에 보인 제안한 방법의 결과는 3GPP 결과보다 넓은 부호화 대역폭을 보이고, 스펙트럼 공백 또한 많이 사라졌음을 알 수 있다. 제안한 방법이 이와 같이 향상된 복원

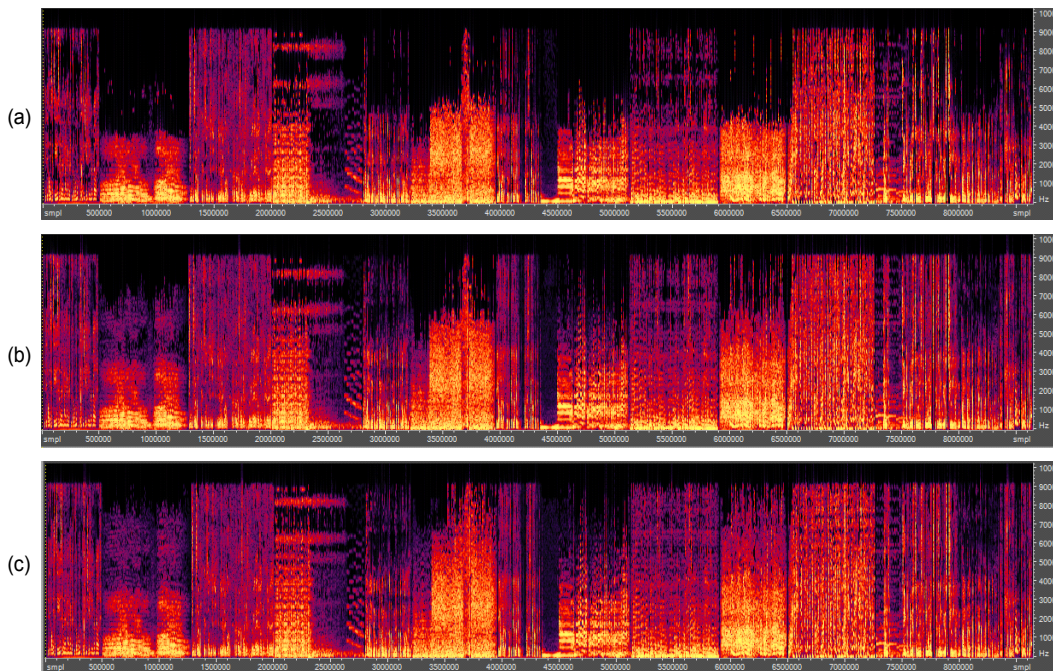


그림 6. (a) 3DA RM9, (b) 3GPP, (c) 제안한 방법으로 all_stereo.wav을 부호화하여 복원한 왼쪽 채널 스펙트럼
Fig. 6. Left channel spectra of the all_stereo.wav reconstructed by (a) 3DA RM9, (b) 3GPP, (c) the proposed method

스펙트럼을 보임에 따라 MUSHRA 테스트에서도 높은 점수를 얻은 것으로 판단된다. 이러한 점들을 종합할 때 본 논문에서 제안한 버퍼 제어 방법은 기존의 RM9이나 3GPP의 버퍼 제어 방법에 비해 비교 우위의 성능을 나타낼 수 있다.

V. 결론

본 논문에서는 3DA 핵심 부호화기의 FD 처리부의 성능 향상을 위해 새로운 버퍼 제어 기법을 제안하였다. 제안한 버퍼 제어 기법은 3GPP의 심리음향 모델과 스펙트럼 공백 회피 기법을 포함한 율-왜곡 제어부를 채택하였다. 그리고, 입력 프레임의 활동성에 따라 비트량을 적절히 할당하기 위해 3GPP 버퍼 제어 방법을 수정하였다. 먼저 목표 PE 값을 낮추고 현재 프레임에 할당할 수 있는 비트량의 최대치를 목표 비트량으로 설정하였다. 그리고, 비트 할당을 채널 단위가 아니라 프레임 전체에 대하여 수행함으로써 좀 더 유연한 비트 할당이 가능하도록 하였다. 그리고, PE 값의 범위를 조정하는 $peMin$ 과 $peMax$ 의 갱신 속도를 낮추고 pe 값과 버퍼 상태에 따라 비트 할당을 수행하는 파라미터들을 조정함으로써, 활동성 높은 프레임에 충분한 비트 할당이 이루어질 수 있도록 하였다. 본 논문에서 제안하는 버퍼 제어 방법은 3DA RM9, 3GPP 버퍼 제어 방법과 그 성능을 비교 평가하였다. 본 논문에서는 3분 길이의 실험 오디오 샘플을 3가지 버퍼 제어 방법으로 부호화한 다음,

8명의 오디오 전문가를 대상으로 한 MUSHRA 테스트로 그 성능을 평가하였다. 성능 평가 결과, 제안한 방법은 RM9 및 3GPP에 비해 평균적으로 약 30 및 13 만큼 높은 점수를 보여, 제안한 방법의 성능이 우수함을 입증할 수 있었다.

본 논문에서는 3DA 핵심부호화기의 FD 처리부의 성능 향상 방안에 대해 살펴보았다. 본 논문에서 제안한 방법은 IGF, SBR, MPEG surround 등과 결합하여 적용하면 비교적 낮은 전송률에서도 매우 높은 성능을 얻을 수 있을 것으로 기대한다.

참고 문헌 (Reference)

- [1] ISO/IEC 13818-7:2003(E), Generic Coding of Moving Pictures and Associated Audio Part 7: Advanced Audio Coding, 2003.
- [2] ISO/IEC 14496-3:2005, Information Technology - Generic Coding of Audiovisual Objects Part 3: Audio, 2005.
- [3] ISO/IEC FDIS 23003-3:2011(E), Information Technology - MPEG Audio Technology - Part 3: Unified Speech and Audio Coding, 2011.
- [4] ISO/IEC FDIS 23008-3:2014(E), Information Technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments - Part 3: 3D Audio, 2014.
- [5] 3GPP TS 26.410 Enhanced AACPlus General Audio Codec, Floating-Point ANSI-C Code (Release 15), June 2018.
- [6] MPEG 3DA Reference Model 9, 2017.
- [7] J. Herre and S. Dick, "Psychoacoustic models for perceptual audio coding-A tutorial review, Appl. Sci., vol. 9, no. 14, 2854, July 2019.
- [8] ITU-R BS.1534-1, Method for the Subjective Assessment of Intermediate Sound Quality(MUSHRA), 2011.

저 자 소 개



김 인 철

- 1985년 : 서울대학교 제어계측공학과(공학사)
- 1987년 : 서울대학교 제어계측공학과(공학석사)
- 1992년 : 서울대학교 제어계측공학과(공학박사)
- 1992년 8월 ~ 1994년 3월 : 대우전자 영상연구소
- 1994년 3월 ~ 1999년 8월 : 한성대학교 정보전산학부 조교수
- 1999년 9월 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학부 교수
- ORCID : <https://orcid.org/0000-0002-7643-8717>
- 주관심분야 : 디지털 신호처리, 실시간 신호처리, 오디오 압축, 멀티미디어 통신 등