



A Study on Quality Improvement of Ultra Multi-channel Audio Coding using Residual Signal

Kwangki Kim*

School of IT Convergence, Korea Nazarene University

ABSTRACT

The conventional ultra multi-channel audio coding (UMAC) based on playback space segmentation divides the playback space into five spaces and separately encodes input signals for each space using the encoding module for each space. Generated 5 down-mix signals for each space are encoded using a 5.1 channel audio coder. Therefore, the UMAC requires only one or two coders, so that the bit rate and complexity are very low, enabling realistic sound service in wired/wireless network, communication, and mobile environments. However, since the UMAC restores the ultra multi-channel signal using only the down-mix signal and the power ratio parameters extracted from the 28 sub-bands, deterioration of the sound quality can not be avoided. In this paper, we propose an enhanced UMAC (E-UMAC) to improve the quality of reconstructed signal by using residual signal which compensates the difference between original signal and reconstructed signal to solve the problem of sound quality deterioration of the UMAC. According to the proposed method, a coding module for each playback space is changed to extract a residual signal and add a process of restoring the input signals using the residual signal. Moreover, the overall coding structure is changed so that one residual signal is extracted for the 2-channel input. For the verification, we coded 10.2 channel audio contents using the implemented E-UMAC and performed the bit-rate measurement and the sound quality evaluation. The effectiveness of the proposed E-UMAC was validated from experimental results that the bit-rate is increased by about 44% compared with the UMAC but the sound quality of the reconstructed signal is greatly improved.

© 2019 KKITS All rights reserved

KEYWORDS: Realistic audio service, Ultra multi-channel audio signals, Ultra multi-channel audio coding, Down-mixing, Spatial parameter, Residual signal

ARTICLE INFO: Received 6 December 2018, Revised 3 January 2019, Accepted 8 February 2019.

*Corresponding author is with School of IT Convergence,
Korea Nazarene University, 48 Wolbong-ro Seobuk-gu

Cheonan-si Chungcheongnam-do, 31172, S.KOREA.
E-mail address: k2kim@kornu.ac.kr

1. 서론

디지털 기술 및 통신 기술의 발달로 방송, 영화, 인터넷 및 개인 미디어 등 다양한 분야에서 오디오 및 비디오 위주의 멀티미디어 콘텐츠의 보급이 급속히 확대되고 있다. 또한, 3DTV와 3D 영화의 보편화와 지상파 UHDTV (Ultra High Definition TV) 방송이 개시됨에 따라 사실감과 현장감을 전달하는 실감미디어에 대한 요구가 커지고 있다 [1]. 이러한 실감미디어는 기존의 일반적인 멀티미디어 콘텐츠에 비해 데이터양이 매우 크기 때문에 유/무선 네트워크나 통신환경에서 실감미디어를 전송하기 위한 비디오 및 오디오 부호화에 대한 연구가 활발히 진행 중이다 [2][3]. 본 논문에서는 유/무선 네트워크나 통신 환경에서 초다채널 오디오 신호에 의한 실감음향 서비스를 위한 부호화 기술을 다루며, 낮은 비트율과 복잡도를 고려하여 개발된 기존의 초다채널 오디오 부호화기[4]의 음질 측면에서의 성능 개선을 위한 방법을 제안한다.

부호화를 거치지 않은 10.2 채널 이상의 초다채널 오디오 신호는 채널 수에 따라 데이터 양이 매우 크게 증가하므로 네트워크나 통신환경을 통해 전송되어 서비스 되는데 많은 제약이 있다. 이를 해결하기 위하여 초다채널 오디오 신호의 부호화에 대한 연구가 진행 중이며, MPEG (Moving Picture Experts Group) 오디오 서브그룹에서는 모노/스테레오 음성/오디오 통합 부호화기인 MPEG-D USAC (Unified Speech and Audio Coding)[5] 을 조합하여 초다채널 오디오 신호를 부호화하는 MPEG-H 3D Audio 표준을 개발하였다 [6]. MPEG-H 3D Audio는 2개의 입력 신호를 하나의 MPEG-D USAC으로 부호화 하므로 10.2 채널의 경우 5개의 부호화기가 필요하여 높은 비트율과 고사양의 시스템이 필요하다는 문제를 지닌다. 이를 해결하고자 낮은 비트율과 복잡도에 초점을 맞춰

서 기존의 5.1 채널 오디오 부호화기인 MPEG Surround[7]나 SSLCC (Sound Source Location Coefficients Coding)[8, 9]를 재활용하는 재생공간 분할 기반의 초다채널 오디오 부호화기가 제안되었다 [4]. 재생공간 분할 기반의 초다채널 오디오 부호화기는 하나 또는 두 개의 오디오 부호화기만을 이용하기 때문에 비트율과 복잡도가 매우 낮아 유/무선 네트워크 및 통신환경과 모바일 환경에서 실감음향 서비스를 가능하게 한다. 그러나, 재생공간 분할 기반의 초다채널 오디오 부호화기는 28개의 서브밴드에서 추출된 파워비 파라미터 만으로 초다채널 신호를 복원하므로 음질의 열화가 크다는 문제를 지닌다. 이를 해결하고자 본 논문에서는 원 신호와 파라미터로 복원한 신호 사이의 차이를 보정해주는 잔차신호를 이용하여 복원된 신호의 음질을 개선하는 방법을 제안한다. 제안된 방법에 따라 기존의 초다채널 오디오 부호화기의 재생공간별 부호화 모듈에 잔차신호를 추출하고 이를 이용하여 복원하는 과정이 추가되며, 잔차신호가 2채널 입력에 대해 1개씩 추출되도록 전체적인 부호화 구조가 변경된다. 최종적으로 제안된 잔차신호 기반의 초다채널 오디오 부호화기를 구현하여 비트율의 측정과 음질 평가를 통해 제안된 방법의 유효성을 확인하도록 한다.

본 논문은 다음과 같이 구성된다. 먼저 2장에서 기존의 초다채널 오디오 부호화기를 설명하고, 3장에서는 제안된 잔차신호 처리를 갖는 초다채널 오디오 부호화기를 설명한다. 4장에서는 실험 결과를 제시하고, 5장에서 결론을 맺도록 한다.

2. 초다채널 오디오 부호화기

2.1 초다채널 오디오 부호화

초다채널 오디오 부호화는 입력 초다채널 신호

를 5개의 공간(Center, Left front, Right front, Left back, Right back)으로 분할하여 수행하며[4], <그림 1>에서 보는 바와 같은 재생공간별 부호화모듈을 각각의 재생공간에 적용한다. 재생공간별 부호화 모듈에서 입력된 공간신호는 다운믹싱 과정을 거쳐 공간을 대표하는 하나의 다운믹스 신호를 생성하며, 재생공간별 신호로의 복원을 위한 파라미터를 추출한다. 재생공간별 다운믹스 신호는 식 (1)을, 파라미터는 식 (2)와 (3)을 이용하여 계산한다.

$$dmx(n) = \sum_{i=1}^N x_i(n), \quad 0 \leq n \leq M-1 \quad (1)$$

$$Pr_i(b) = \frac{P_i(b)}{P_{\max}(b)}, \quad \begin{cases} 1 \leq b \leq L \\ 1 \leq i \leq N \end{cases} \quad (2)$$

$$P_i(b) = \sum_{k=A_{b-1}}^{A_b-1} [X_i(k)]^2 \quad (3)$$

식 (1)에서 n은 시간 인덱스, i는 채널 인덱스를 나타내며, N은 재생공간별 입력되는 채널 개수, M은 프레임 길이로 2048의 값을 갖는다. 식 (2)에서 P_{max}는 입력되는 신호 중 최대파워 값을 나타내며, b는 서브밴드 인덱스, L은 서브밴드 수로 28의 값을 갖는다. 식 (3)에서 k는 주파수 인덱스를 나타내

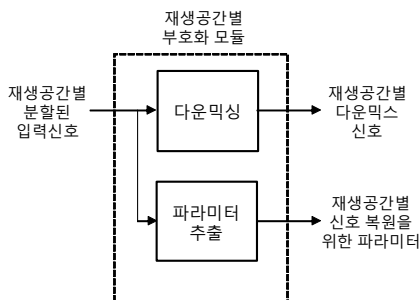


그림 1. 재생공간별 부호화 모듈.
Figure 1. Encoding module for each playback space

며, X_i(k)는 i번째 입력 신호를 주파수 영역으로 변환한 것이다. 또한, A_{b-1}과 A_b-1은 서브밴드의 주파수 영역에서의 시작과 끝 지점을 나타내며, 인간의 청각특성을 반영하여 28개의 서브밴드는 <표 1>의 값을 사용한다 [10].

표 1. 서브밴드의 구분 기준 (DFT 길이: 2048, 샘플링 주파수: 44.1 kHz)
Table1. Classification criteria of subband (DFT length: 2048, sampling frequency: 44.1 kHz)

A ₀	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇
0	3	7	11	15	19	23	27
A ₅	A ₉	A ₁₀	A ₁₁	A ₁₂	A ₁₃	A ₁₄	A ₁₅
31	39	47	55	63	79	95	111
A ₁₆	A ₁₇	A ₁₈	A ₁₉	A ₂₀	A ₂₁	A ₂₂	A ₂₃
127	159	191	223	255	287	319	367
A ₂₄	A ₂₅	A ₂₆	A ₂₇	A ₂₈	-	-	-
415	479	559	655	1025			

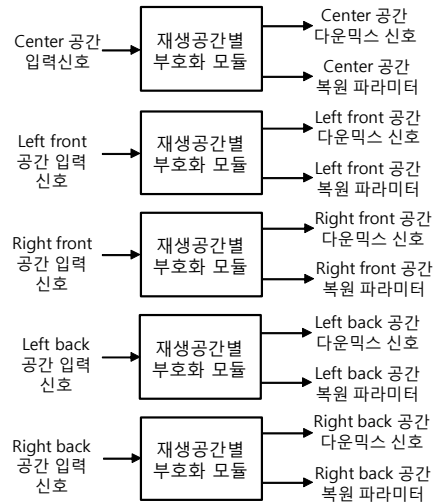


그림 2. 재생공간별 부호화 모듈을 활용한 초다채널 오디오 부호화 구조.

Figure 2. Ultra multi-channel audio encoding structure using coding module for each playback space

이러한 재생공간별 부호화 모듈은 <그림 2>처럼 각 재생공간별로 적용되어 각 재생공간을 대표하는 5개의 다운믹스 신호와 재생공간별 신호 복원

을 위한 파라미터를 추출한다. 재생공간별 부호화 과정을 통해 생성된 다운믹스 신호들은 <그림 3>처럼 기존의 MPEG Surround나 SSLCC와 같은 다채널 오디오 부호화기[7-9]를 통해 최종 스테레오 다운믹스 신호와 다채널 신호 복원을 위한 파라미터를 생성한다.

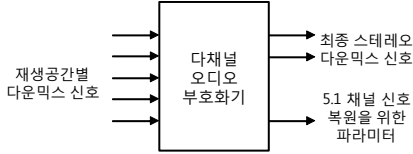


그림 3. 재생공간별 다운믹스 신호의 부호화.
Figure 3. Encoding of down-mix signal for each playback space.

2.2 초다채널 오디오 복호화

초다채널 신호의 복원은 부호화의 역과정으로 진행된다. 먼저, <그림 4>에서 보는 바와 같이 다채널 오디오 복호화기에서 스테레오 다운믹스 신호와 다채널 신호 복원을 위한 파라미터를 이용하여 다채널 오디오 신호 (공간별 다운믹스 신호)를 생성한다 [4]. 생성된 다채널 재생공간별 다운믹스 신호는 <그림 5>와 <그림 6>의 재생공간별 복호화 모듈에 입력되어 재생공간별 신호를 생성하게 된다. 재생공간별 복호화는 수식 (4)를 이용하여 수행한다.

$$\hat{X}_i(k) = G_i(b) \cdot DMX(k), \begin{cases} 1 \leq i \leq N \\ 1 \leq b \leq L \\ A_{b-1} \leq k \leq A_b - 1 \end{cases} \quad (4)$$

$$G_i(b) = \sqrt{\frac{Pr_i(b)}{\sum_{j=1}^N Pr_j(b)}} \quad (5)$$

수식 (4)에서 $\hat{X}_i(k)$ 는 주파수 영역에서 복원된 i 번째 채널 신호이며, $DMX(k)$ 는 주파수 영역으로

변환된 다운믹스 신호이다. 수식 (5)의 $G_i(b)$ 는 서브밴드 b 에서 i 번째 채널 신호의 이득을 나타내는 것으로 주파수 영역으로 변환된 다운믹스 신호에 이를 곱해서 원래의 초다채널 신호를 복원할 수 있다.

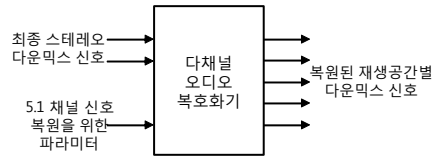


그림 4. 재생공간별 다운믹스 신호의 복호화.
Figure 4. Decoding of down-mix signal for each playback space.

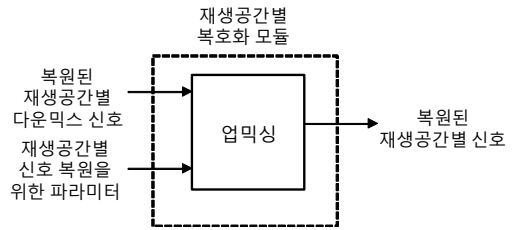


그림 5. 재생공간별 복호화 모듈.
Figure 5. Decoding module for each playback space

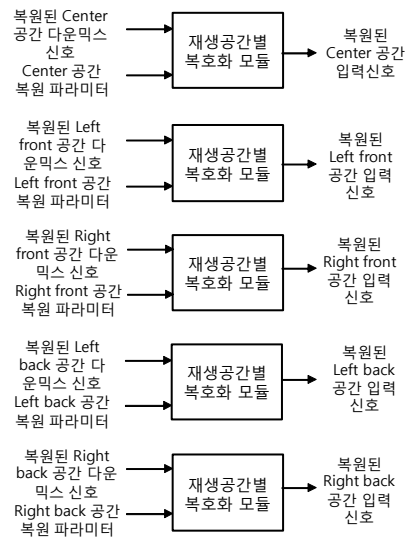


그림 6. 재생공간별 부호화 모듈을 활용한 초다채널 오디오 부호화 구조.

Figure 6. Ultra multi-channel audio decoding structure using coding module for each playback space

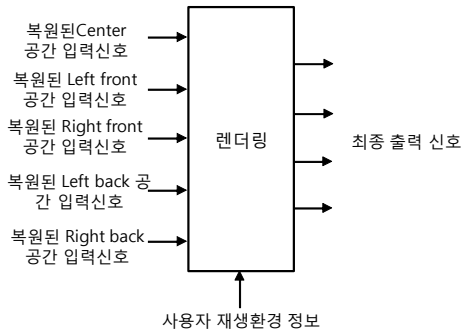


그림 7. 사용자 재생환경에 따른 렌더링을 통한 최종 신호 출력
Figure 7. Final signal output through rendering according to user playback environment

이러한 복호화 과정을 거쳐 생성된 초다채널 오디오 신호는 <그림 7>에서 보이는 렌더링 과정을 통해 사용자의 재생환경에 따른 최종 출력 신호를 생성하며, 스피커 재생환경과 헤드폰 재생환경에 따라 매트릭스 기반의 렌더링과 바이노럴 렌더링을 적용한다 [11][12].

3. 잔차 신호를 활용한 초다채널 오디오 부호화기 개선

3.1 개요

2장에서 설명한 초다채널 오디오 부호화기의 재생공간별 부호화 모듈은 다운믹스 신호와 각 채널 신호의 파워비만을 활용하기 때문에 재생공간별 오디오 신호가 부호화와 복호화 과정에서 음질의 열화가 심하게 발생한다. 특히, 각 채널 신호의 파워비 파라미터는 주파수 영역을 28개의 서브밴드로 나눠서 추출하기 때문에 낮은 주파수 해상도를 지녀 음질 열화의 주요 요인이 된다. 이를 해결하기 위해서는 현재의 28개의 서브밴드의 수를 높여 주파수 해상도를 높이는 방법과 각 재생공간별 원 신호와 복원된 신호 사이의 잔차신호[13]를 이용하여 복원된 신호를 최대한 원 신호에 가깝게 만든

는 방법을 고려할 수 있다. 본 논문에서는 잔차신호를 이용하여 재생공간별 부호화 모듈을 개선함으로써 초다채널 오디오 부호화기의 성능을 개선하도록 한다.

3.2 개선된 초다채널 오디오 부호화

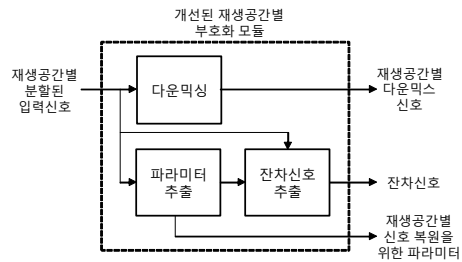


그림 8. 개선된 재생공간별 부호화 모듈.
Figure 8. Enhanced encoding module for each playback space

<그림 8>은 잔차신호의 추출이 추가된 재생공간별 부호화 모듈로 잔차신호는 입력신호와 추출된 파라미터를 이용하여 수식 (6)과 같이 추출한다 [14].

$$R(k) = \frac{G_2(b) \cdot X_1(k)}{G_1(b) + G_2(b)} - \frac{G_1(b) \cdot X_2(k)}{G_1(b) + G_2(b)}, \begin{cases} 1 \leq b \leq L \\ A_{b-1} \leq k \leq A_b - 1 \end{cases} \quad (6)$$

수식 (6)에서 $R(k)$ 는 잔차신호, $X_1(k)$ 와 $X_2(k)$ 는 입력신호, $G_1(b)$ 와 $G_2(b)$ 는 추출된 파라미터와 수식 (5)를 이용하여 계산된 입력신호의 이득을 나타낸다. 한편, 잔차신호는 수식 (6)과 같이 2개의 입력 신호와 이들의 복원을 위한 파라미터를 이용하여 추출하므로 이에 따라 재생공간별 부호화 모듈은 입력 신호를 2개씩 처리하도록 조정된다. <그림 9>, <그림 10>, <그림 11>에 보이는 바와 같이 재생공간별 입력 신호의 개수가 증가함에 따라 재생공간별 부호화 모듈을 조합하여 재생공간별 오디오 신호의 부호화를 수행하며, 입력신호가 5개 이상의

경우에도 마찬가지로 재생공간별 부호화 모듈을 확장하여 부호화하도록 한다.

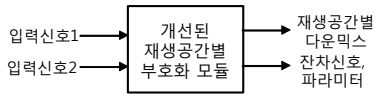


그림 9. 2 채널 입력 신호의 부호화.
Figure 9. Encoding of 2-channel input signals.

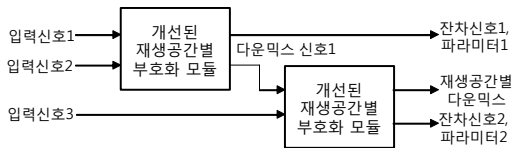


그림 10. 3 채널 입력 신호의 부호화.
Figure 10. Encoding of 3-channel input signals.

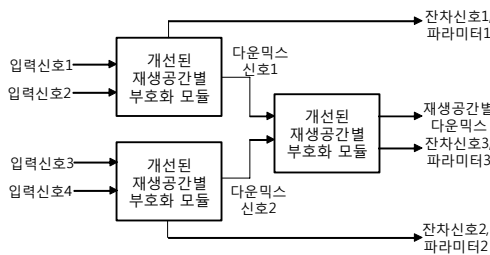


그림 11. 4 채널 입력 신호의 부호화.
Figure 11. Encoding of 4-channel input signals.

3.3 개선된 초다채널 오디오 복호화

잔차신호를 활용하여 개선된 재생공간별 부호화 모듈에 따라 복호화 모듈도 <그림 12>에서 보는 바와 같이 수정되며, 재생공간별 다운믹스 신호, 잔차신호, 파라미터를 활용한 재생공간별 신호의 복원은 수식 (7)을 이용하여 수행된다.

$$\begin{cases} \hat{X}_1(k) = DMX(k) \cdot G_1(b) + \hat{R}(k) \\ \hat{X}_2(k) = DMX(k) \cdot G_2(b) - \hat{R}(k) \end{cases} \begin{cases} 1 \leq b \leq L \\ A_{b-1} \leq k \leq A_b - 1 \end{cases} \quad (7)$$

수식 (7)에서 $\hat{X}_1(k)$ 와 $\hat{X}_2(k)$ 는 복원된 신호를

나타내며, $\hat{R}(k)$ 는 부호화/복호화 과정을 거친 잔차신호이다. 재생공간별 입력 신호의 개수에 따라 재생공간별 부호화를 수행한 것처럼 재생공간별 복호화는 재생공간별 복호화 모듈을 조합하여 <그림 13>, <그림 14>, <그림 15>와 같이 재생공간별 복호화 모듈을 확장하여 신호를 복원하도록 한다.

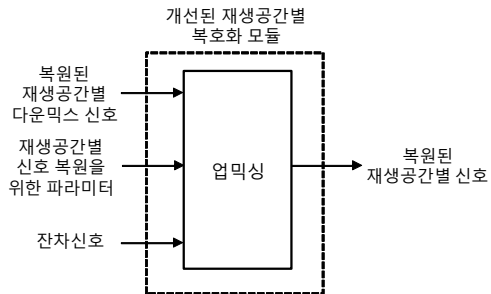


그림 12. 개선된 재생공간별 복호화 모듈.
Figure 12. Enhanced decoding module for each playback space

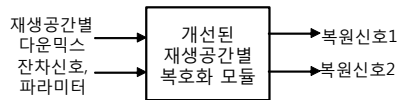


그림 13. 2채널 신호의 복호화.
Figure 13. Decoding of 2-channel signals.

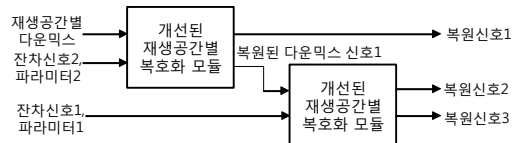


그림 14. 3 채널 신호의 복호화.
Figure 14. Decoding of 3-channel signals.

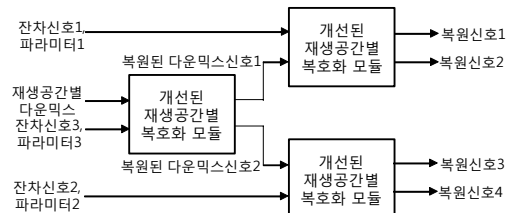


그림 15. 4 채널 신호의 복호화.
Figure 15. Decoding of 4-channel signals.

3.4 잔차신호의 부호화

잔차신호는 수식 (6)에서 보는 바와 같이 전체 주파수 대역에서 추출하므로 그 자체로 또 다른 오디오 신호이며, 별도로 부호화 및 복호화 과정을 거친다. 잔차신호를 일반적인 오디오 신호처럼 부호화를 했을 경우, 높은 비트율을 필요로 하기 때문에 초다채널 오디오 부호화기의 음질을 개선한다고 할지라도 비트율 측면에서의 성능을 크게 저하 시키는 문제를 지닌다. 이에 따라 인간의 청각 특성이 저주파에 민감하다는 사실에 기반하여 잔차신호의 저주파 대역인 0-5.5 kHz에 해당하는 신호만 사용하도록 한다 [14]. 잔차신호의 부호화는 TCX (Transform Coded Excitation) 기반의 오디오 부호화기 [14][15]를 이용하여 수행하며, 채널 당 15 kbps를 할당함으로써 잔차신호에 의한 비트율 증가를 최소화 하고 기존의 초다채널 오디오 부호화기의 음질을 향상시키도록 하였다.

4. 실험 및 결과

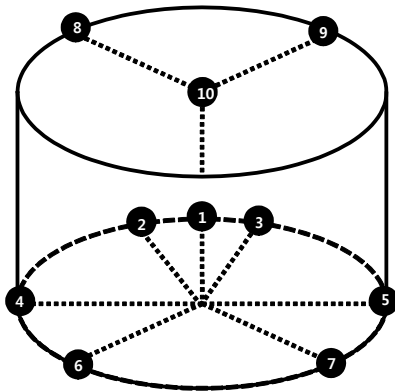


그림 16. 10.2 채널 레이아웃
Figure 16. 10.2 channel configuration

잔차신호를 활용하여 개선된 초다채널 오디오 부호화기의 성능을 확인하기 위하여 10.2 채널 오

디오 콘텐츠 3개를 이용하여 부호화를 수행한 후, 비트율과 음질을 확인하였다. 오디오 콘텐츠는 20 초의 지속시간과 44.1 kHz 샘플링 주파수, 16 비트 양자화를 갖으며, <그림 16>과 <표 2>에서 보는 바와 같이 각 채널 공간별 2개의 오디오 신호가 할당된다.

표 2. 10.2 채널 재생공간 분할
Table 2. Space segmentation of 10.2 channel configuration

분할 공간	할당 채널
Center	1, 10
Left front	2, 8
Left back	4, 6
Right front	3, 9
Right back	5, 7

표 3. 비트율 결과

Table 3. Bitrate result

모드	비트율 평균	증가율
기존의 초다채널 오디오 부호화기	170.15 kbps (다운믹스 140 kbps+채널당 약 3 kbps)	44.08 %
잔차신호를 활용하여 개선된 초다채널 오디오 부호화기	245.15 kbps (다운믹스 140 kbps+채널당 10.5 kbps)	

<표 3>은 기존의 초다채널 오디오 부호화기와 개선된 초다채널 오디오 부호화기에 의한 10.2 채널 오디오 콘텐츠의 부호화 비트율을 보여준다. 잔차신호를 제외한 비트율의 증가는 없으며, 2개의 입력신호 별로 하나의 잔차신호를 추출하므로 5개의 잔차신호가 생성되고 이를 15 kbps로 부호화하기 때문에 75 kbps의 비트율이 증가한다. <그림 17>은 <표 4>의 시스템과 6명의 평가자를 대상으로 수행한 MUSHRA [16] 청취평가 결과를 보여준다. 본 실험에서는 10.2 채널 재생 시스템의 구현상의 문제로 재생공간별 2채널 원 신호와 복원 신호의 음질을 비교하여 잔차신호에 의한 음질 개선을 확

인하였다. 청취평가 결과 잔차신호를 활용하여 개선된 초다채널 오디오 부호화기에 의해 복원된 신호가 기존의 초다채널 오디오 부호화기에 의해 복원된 신호보다 크게 개선된 음질을 보인다. 이는 잔차신호에 의해 비트율이 약 44% 증가했음에도 음질개선 측면에서 매우 효율적임을 보여주어 제안된 잔차신호를 갖는 초다채널 부호화기의 유효성을 확인할 수 있다.

표 4. 실험 대상
Table 4. System under test

구분	설명
Reference	원 신호
Original	기존의 초다채널 오디오 부호화기에 의해 복원된 신호
Enhanced	잔차신호를 활용하여 개선된 초다채널 오디오 부호화기에 의해 복원된 신호

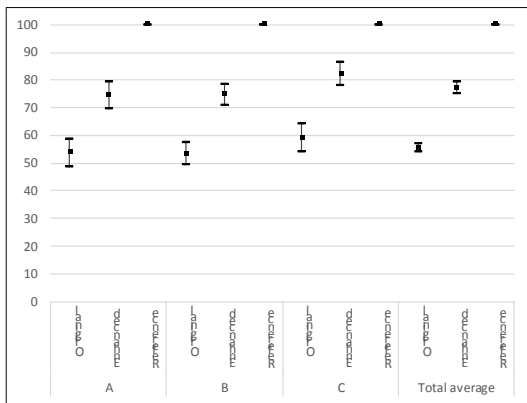


그림 17. 청취평가 결과
Figure 17. Listening test results.

5. 결론

본 논문에서는 기존의 초다채널 오디오 부호화기의 음질 측면에서의 성능을 개선하기 위하여 잔차신호를 갖는 초다채널 오디오 부호화 방법을 제안하였다. 제안된 방법은 각 재생공간별로 부호화

모듈을 이용하여 부호화하는 기존의 초다채널 오디오 부호화기에 각 재생공간별로 잔차신호를 추가로 추출하여 파라미터만으로 복원된 신호가 원신호에 가까워지도록 함으로써 음질을 개선하였다. 잔차신호의 추가에 따라 기존의 재생공간별 부호화 모듈은 2 채널 입력만을 받도록 수정되었으며, 입력 신호의 수에 따라서 재생공간별 부호화 모듈을 조합하여 초다채널 오디오 신호를 부호화하였다. 추출된 잔차신호는 TCX 기반의 오디오 부호화기를 이용하여 압축/복원하며, 채널 당 15 kbps의 비트율이 소모되도록 하였다. 실험을 통해 제안된 잔차신호를 갖는 초다채널 오디오 부호화기가 기존의 초다채널 오디오 부호화기에 비해 약 40%의 비트율 증가를 보이는데 반해 음질을 크게 개선하여 제안된 방법의 유효성을 확인할 수 있었다. 향후 과제로 제안된 방법에 의한 전체적인 음질의 개선 정도를 확인하기 위하여 10.2 채널 재생 시스템에서의 청취평가가 필요하며, 서브밴드 수의 조절을 통해 잔차신호보다 비트율의 증가는 최소화하면서 음질을 개선할 수 있는 방법에 대한 연구가 필요하다.

References

- [1] S. Cho, D. Jeon, and J. Choi. *Status on UHDTV technology and standadization*, TTA Journal 140, pp. 49-54, 2012.
- [2] J. Shin, *Status and prospect of coding technology for UHDTV broadcasting*, Broadcasting and Media, Vol. 19, No. 2, pp. 69-79, 2014.
- [3] J. Seo, J. Yoo, K. Kang, and S. Jang, *Status of audio standardization for realistic broadcasting*, Broadcasting and Media, Vol. 19, No. 1, pp. 37-47, 2014.

- [4] K. Kim, *A study on ultra multi-channel audio coding for realistic audio services*, Journal of Knowledge Information Technology and Systems, Vol. 13, No. 1, pp 101-120, 2018.
- [5] ISO/IEC 23003-3:2012, *Information technology-MPEG audio technologies-Part 3: Unified speech and audio coding*, 2012.
- [6] ISO/IEC 23008-3:2015, *MPEG-H (High efficiency coding and media deliver in heterogeneous environments), Part 3: 3D Audio*, 2015.
- [7] ISO/IEC 23003-1:2007, *Information technology-MPEG audio technologies-Part 1: MPEG surround*, 2007.
- [8] H. Moon, J. Seo, S. Beack, and K. Sung, *A multi-channel audio compression method with virtual source location information for MPEG-4 SAC*, IEEE Transactions on Consumer Electronics, Vol. 51, No. 4, pp. 1253-1259, 2005.
- [9] S. Beack, J. Seo, H. Moon, K. Kang, and M. Hahn, *Angle-based virtual source location representation for spatial audio coding*. ETRI Journal, Vol. 28, No. 2, pp. 219-222, 2006.
- [10] F. Baumgarte, and C. Faller, *Binaural cue coding-part I: psychoacoustic fundamentals and design principles*, IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 6, pp. 509-519, 2003.
- [11] K. Kim, *Implementation of binaural decoding in spatial cue based multi-channel audio coding*, Journal of Knowledge Information Technology and Systems, Vol. 9, No. 1, pp. 85-94, 2014.
- [12] K. Kim, *A study on sound scene control of binaural sound in multi-channel audio coding*, Journal of Knowledge Information Technology and Systems, Vol. 12, No. 1, pp. 177-185, 2017.
- [13] J. Herre, H. Purnhagen, J. Breebard, C. Faller, S. Disch, K. Kjørling, E. Schuijers, J. Hilpert, and F. Myburg, *The reference model architecture for MPEG spatial audio coding*, in Proc. 118th AES Convention, Barcelona, Spain, 2005.
- [14] K. Kim, S. Beack, J. Seo, K. Kang, and M. Hahn, *Spatial audio object coding with two-step coding structure for interactive audio service*, IEEE Transactions on Multimedia, Vol. 13, No. 6, pp. 1208-1216, 2011.
- [15] R. Lefebvre, R. Salami, C. Laflamme, and J.-P. Adoul, *High quality coding of wideband audio signals using transform coded excitation (TCX)*, in Proc. ICASSP'94, pp. 1.193-1.196, Apr. 1994.
- [16] ITU-R Recommendation, *Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)*, ITU, BS 1543-1. Geneva, Switzerland, 2001.

잔차신호를 활용한 초다채널 오디오 부호화기의 음질개선에 관한 연구

김광기

나사렛대학교 IT융합학부 교수

요 약

재생공간 분할 기반의 초다채널 오디오 부호화기는 재생공간을 5개 공간으로 나눠 각 공간별로 부호화한 후, 생성된 다운믹스 신호를 5.1 채널 오디오 부호화기를 이용하여 부호화하기 때문에 하나 또는 두 개의 부호화기만 필요하여 비트율과 복잡도가 매우 낮아 유/무선 네트워크 및 통신환경과 모바일 환경에서

실감음향 서비스를 가능하게 한다. 그러나, 재생공간 분할 기반의 초다채널 오디오 부호화기는 다운믹스 신호와 28개의 서브밴드에서 추출된 파워비 파라미터만으로 초다채널 신호를 복원하기 때문에 음질의 열화를 피할 수 없다. 초다채널 오디오 부호화기의 음질 열화 문제를 해결하기 위해 원 신호와 복원된 신호 사이의 차이를 보정해주는 잔차신호를 이용하여 복원된 신호의 음질을 개선하는 방법을 제안한다. 제안된 방법에 따라 기존의 재생공간별 부호화 모듈은 잔차신호를 추출하고 이를 이용하여 복원하는 과정이 추가되며, 2채널 입력에 대해 하나의 잔차신호가 추출되도록 전체적인 부호화 구조가 변경된다. 검증을 위해 구현된 잔차신호 기반의 초다채널 오디오 부호화기를 이용하여 10.2 채널 오디오 콘텐츠에 대해 부호화를 수행하였으며, 비트율 측정과 음질평가를 수행하였다. 실험결과로부터 비트율은 기존의 부호화기에 비해 약 44% 증가하였으나 음질이 크게 개선되어 제안된 방법의 유효성을 확인할 수 있었다.

South Korea, in 2004 and 2011, respectively. He worked as a researcher in Samsung DMC R&D center in 2012. Currently, He is an assistant professor in Korea Nazarene University, Chon-an, Korea. His research interests include multi-channel/-object audio coding, spatial audio, 3D sound, and their applications. He is a life member of the KKITS.

E-mail address: k2kim@kornu.ac.kr

Acknowledgments

본 논문은 2018년도 나사렛대학교 교비학술연구조성비 지원에 의해 연구되었으며, 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2017R1D1A3B03034951). 또한, 본 연구는 한국연구재단을 통해 과학기술정보통신부의 기초연구실지원사업으로부터 지원받아 수행됨(NRF-2018R1A4A1025559).



Kwangki Kim received the B.S. degree in electronic engineering from Korea Aviation University, Koyang, South Korea, in 2002, the M.S. degree and the Ph.D. degree in department of Information and Communications Engineering at Korea Advanced Institute of Science and Technology, Daejeon,