Coding Structure for Ultra Multi-channel Audio Signals

Kwangki Kim
School of IT convergence
Korea Nazarene University
Cheonan, Korea
k2kim@kornu.ac.kr

Kiho Kim
Dept. of digital contents
Korea Nazarene University
Cheonan, Korea
isk01259@gmail.com

Seungmin Oh
Dept. of digital contents
Korea Nazarene University
Cheonan, Korea
osm5252kr@gmail.com

Jinsul Kim
Dept. of electronics and computer engineering
Chonnam National University
Kwangju, Korea
jsworld@jnu.ac.kr

Abstract—**In this paper, we proposed an ultra multi-channel audio coding scheme to enable realistic audio service in wired / wireless network or communication system. The proposed coding scheme comes from considering the use of a conventional 5.1 channel reproduction system and a multi-channel audio coder. Since the proposed method uses only one or two existing coders, it has low complexity and low bit rate. So, the proposed ultra multi-channel audio coding scheme can be used for realistic audio service through wired or wireless network or communication system. It has been confirmed that the proposed ultra multi-channel coding scheme with a simple implementation has about 200 kbps to code a 10.2 channel audio signal.**

Keywords— coding structure, realistic audio service, ultra multi-channel audio, ultra multi-channel audio coding

I. Introduction

With the advent of ultra high definition (UHD) video such as 3D movies and UHD broadcasting, there are a growing interest and many researches in realistic audio technologies that reproduce realistic sounds in the real world through the conventional playback system. The representatives of the realistic audio technologies are the ultra multi-channel playback system from 10.2 channel to 22.2 channel or more, the flexible rendering for compensating the mismatch between the number of channel of the transmitted audio signals and the user's reproduction environment, the binaural rendering for generating 3D sound with stereo headphone, and the audio coding for the realistic audio service in the wired/wireless network or communication systems [1-4]. Among the realistic audio technologies, we focused on the audio coding scheme to make the realistic audio service possible through the wired/wireless network or communication systems.

Since the data rate of the ultra multi-channel audio signals greatly increases in proportion to the number of channels, the realistic audio service using the ultra multi-channel audio signals may be impossible through the wired/wireless network or communication systems when the signals are directly delivered without any compression. To solve this problem, studies on the coding schemes related to compression and reconstruction of the ultra multi-channel audio signals have been carried out to efficiently handle a multi-channel audio signal at a level of reusing the coders used for coding the mono/stereo or the multi-channel signal. Recently, MPEG has developed the standardization of the MPEG-H 3D Audio that efficiently encodes and decodes the ultra multi-channel audio signals from 10.2 to 22.2 channels or more [5, 6]. Based on the MPEG-D USAC, which is a unified speech and audio coder that has already been standardized in MPEG, the MPEG-H 3D Audio sorts the ultra multi-channel audio signals by two channel signals and codes each two channel signals using MPEG-D USAC [7, 8]. Since the existing stereo audio coder is recycled, 11 USACs are required for 22.2 channels at least, which results in high bit rate and high computational complexity. In this paper, we proposed an ultra multi-channel audio coding structure that adopts the existing multi-channel coder such as MPEG Surround and sound source location coefficient coding (SSLCC) [9-12]. In the proposed method, we only used one or two existing coders and additional parameters for minimizing the complexity and the bit-rate.

II. The Proposed Coding Structure for Ultra Multi-channel Audio Signals

Fig. 1 shows the proposed ultra multi-channel audio coding structure for reusing an existing multi-channel audio coder. Since the multi-channel audio coders such as MPEG Surround and SSLCC have been developed and they are able to efficiently handle 5.1 channel signals, we reuse the multi-channel audio coders in the proposed ultra multi-channel audio coding. The input ultra multi-channel audio signal is down-mixed into a 5.1-channel audio signal and each channel information such as power, correlation, etc. is extracted and transmitted to a decoding side to recover the ultra multi-channel signal. The generated 5.1 channel audio signal is compressed as a stereo down-mix signal and the side information using a conventional multi-channel audio coder. Finally, one stereo down-mix signal, the channel information for recovering ultra multi-channel signal, and the additional side information for reconstructing the multi-channel signal are generated through the whole encoding process of the proposed ultra multi-channel audio coder. The decoding process is performed by following the encoding process in the reverse order, and the final output signal is generated by rendering the

978-1-5386-3679-4/18/$31.00 ©2018 IEEE

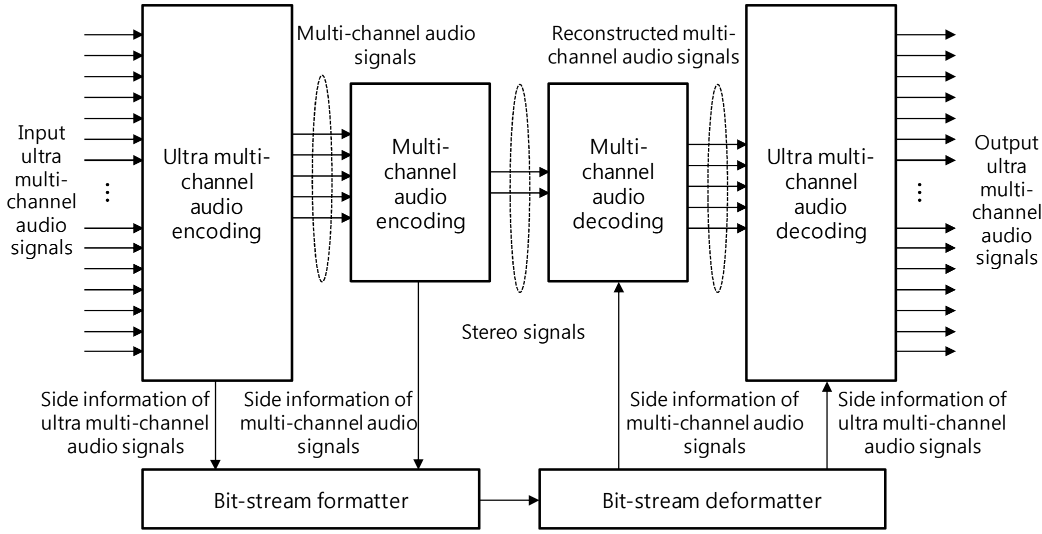

Fig. 1. Overall structure of the proposed ultra multi-channel audio coding

recovered ultra multi-channel signal according to the user's reproduction environment. To encode the ultra multi-channel audio signal, the ultra multi-channel reproduction system is divided into five spaces (center, left front, left back, right front and right back) as shown in Fig. 2. All signals allocated to each space is down-mixed and channel information such as channel power ratio and correlation is extracted for restoring the channel signal corresponding to the space. In other words, considering the 5.1 channel playback system widely used in general households, 10.2 channel or more multi-channel audio signals are down-mixed into 5.1 channels based on the divided five spaces and the parameters are extracted for recovering the ultra multi-channel audio signal with the down-mix signal. The 5.1 channel down-mix signal is encoded using a conventional multi-channel audio coder as needed, and it is compressed and restored with a stereo down-mix signal and another parameter.

For an explanation, we give a simple example of coding in the left front space as shown in Fig.3. When there are four channel signals L_1, L_2, L_3, and L_4 in the left front space, one down-mix signal is generated by adding all four signals as in (1).

$$Lf(k) = L_1(k) + L_2(k) + L_3(k) + L_4(k) \; for \;\; 0 \le k \le N-1 \quad (1)$$

where Lf is the down-mix signal to be considered as the left front signal of 5.1 channel configuration and k is the frequency index. The spatial parameters are calculated by using the power ratio of each channel as in (2).

$$PR_i(b) = \frac{P_i(b)}{P_{\max}(b)} \; for \; 1 \le i \le 4, \; 1 \le b \le 28 \qquad (2)$$

where $P_i(b) = \sum_{k=A_{b-1}}^{A_b - 1} \left[L_i(k)\right]^2$ is the power of the i[th] channel and $P_{\max}(b)$ is the maximum power among all channels. $PR_i(b)$ is the power ratio of the i[th] channel and b is the sub-band index. Here, the sub-band index b is the spectral region to contain several spectral components and 28 sub-bands are used to reflect the human hearing's characteristics [13]. Furthermore, A_{b-1} and $A_b - 1$ are the beginning and end frequency indexes of the b[th] sub-band, respectively. The reconstruction of the original signals in the left front space using the down-mix signal and the parameters can be easily performed by using the following (3).

$$\hat{L}_i(k) = Lf(k) \times \frac{PR_i(b)}{\sum_{j=1}^{4} PR_j(b)} \; for \; \begin{cases} 1 \le i \le 4 \\ 1 \le b \le 28 \\ A_{b-1} \le k \le A_b - 1 \end{cases} \qquad (3)$$

where $\hat{L}_i(k)$ is the recovered signal of the i[th] channel.

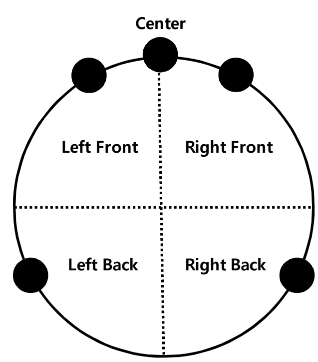

Fig. 2. Space segmentation of the ultra multi-channel audio reproduction system.

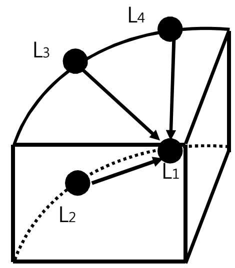

Fig. 3. An example of coding in the left front space.

III. SIMPLE IMPLEMENTATION AND DISCUSSION

We implemented a simple ultra multi-channel audio coder adopting the down-mixing and the power ratio of the audio signal and confirmed the feasibility of the proposed ultra multi-channel audio coder. Ten channel audio signals except for the LFE channel of 10.2 channel audio signals are coded. For the channel configuration in Fig. 4, we applied the playback space segmentation as shown in Fig. 2. Table 1 shows the allocated channel according to the proposed playback space segmentation. The input ultra multi-channel signals are divided into five spaces and they are encoded by the proposed ultra multi-channel audio coder. Through the encoding process, five channel down-mix signals are generated and the parameters for the reconstruction of the original signals are extracted. Here, we assumed that the down-mix signals are encoded using a conventional multi-channel audio encoder. Three 10.2 channel audio contents were used for the experiment and the average bit rate of the extracted parameter, i.e. power ratio, was about 48kbps. Since the bit-rate for the 5 channel down-mix signal is approximately 140 kbps or less, total bit-rate for the 10.2 channel audio signals by the prototype of the proposed ultra multi-channel audio coder is about 200 kbps or less. In this experiment, only the bit rate is investigated through the simple implementation of the ultra multi-channel audio coder. Although the sound quality evaluation of the restored signal is not performed yet, it is convinced that the sound quality degradation is not avoidable due to the down-mixing process, the parameter extraction in the sub-band, a simple usage of the parameter, i.e. only power ratio, and so on.

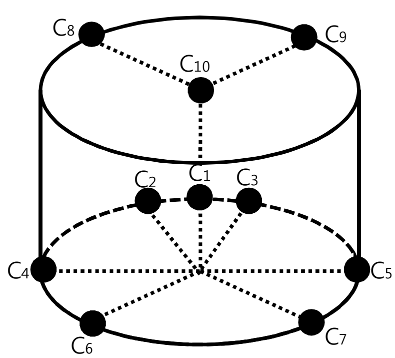

Fig. 4. 10.2 channel configuration

TABLE I. ALLOCATED CHANNEL ACCORDING TO SPACE SEGMENTATION

Segmented channel	Allocated channel
Center	C1, C10
Left front	C2, C8
Left back	C4, C6
Right front	C3, C9
Right back	C5, C7

IV. CONCLUSION

In this paper, we proposed the ultra multi-channel audio coding structure to make the realistic audio service possible in the wired/wireless network or communication systems. The proposed coding structure is originated by considering the conventional 5.1 channel playback system and the usage of the multi-channel audio coder. Since the proposed method only uses one or two existing coders and has rather low complexity and low bit-rate, it can be used for the realistic audio service through the wired/wireless network or communication systems. From the simple implementation, it was confirmed that the proposed ultra multi-channel coding structure has about 200 kbps for coding the 10.2 channel audio signals. As future works, the subjective listening test should be performed to check the sound quality of the proposed coding structure at first and various audio techniques should be investigated and adopted to improve the performance of the ultra multi-channel audio coder.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A3B03034951). Also, this research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2017-2016-0-00314) supervised by the IITP(Institute for Information & communications Technology Promotion).

REFERENCES

[1] Tomlinson Holman, “The History and Future of DSPs in Consumer Audio Equipment — Part II: Emerging Areas and the Future,” ICCE2008.

[2] K. Hamasaki, S. Komiyama, K. Hiyama, H. Okubo, “5.1 and 22.2 Multichannel Sound Productions Using an Integrated Surround Sound Panning System,” 2005 NAB BEC Proceedings, April 2005.

[3] Young Woo Lee, Sunmin Kim, Hyun Jo, Youngjin Park, Yongje Kim, “Virtual Height Speaker Rendering for Samsung 10.2-Channel Vertical Surround System,” 131st AES Convention,Oct. 2011.

[4] Daeyoung Jang, Jeongil Seo, Taejin Lee, Kyeongok Kang, “Present and Future of UHD Sound Technology”, Korea Society Broadcast Engineers Magazine Vol. 17, no. 5, pp.47-59, 2012. 10.

[5] Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J., “MPEG-H 3D Audio - The New Standard for Coding of Immersive Spatial Audio,” IEEE J. Selected Topics in Signal Processing, vol. 9, no. 5, Aug. 2015, pp. 770-779.

[6] ISO/IEC 23008-3:2015, MPEG-H (High efficiency coding and media deliver in heterogeneous environments), Part 3: 3D Audio, 2015.

[7] Neuendorf, M., Multrus, M., Rettelbach, N., Fuchs, G., Robilliard, J., Lecomte, J., & Lefebvre, R., "The ISO/MPEG unified speech and audio coding standard—consistent high quality for all content types and at all

bit rates." Journal of the Audio Engineering Society 61.12 (2013): 956-977.

[8] ISO/IEC 23003-3:2012, "Information Technology – MPEG Audio Technologies – Part 3: Unified Speech and Audio Coding

[9] J. Herre, H. Purnhagen, and J. Breebard, “The reference model architechture for MPEG spatial audio coding,” in Proceedings of the 118th AES Convention, Barcelona, Spain, 2005.

[10] ISO/IEC 23003-1:2007, "Information Technology – MPEG Audio Technologies – Part 1: MPEG Surround"

[11] H. Moon, J. Seo, S. Beack & K. Sung, “A multi-channel audio compression method with virtual source location information for MPEG-4 SAC,” Consumer Electronics, IEEE Transactions on. 01/12/2005; 51(4):1253- 1259.

[12] S. Beack, J. Seo, H. Moon, K. Kang, M. Hahn, “Angle-Based Virtual Source Location Representation for Spatial Audio Coding,” ETRI Journal, Vol. 28, No. 2, pp. 219-222, April 2006.

[13] C. Faller and R. Baumgarte, “Binaural Cue Coding–Part II: Schemes and Application,” IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.