Mastering Signal Processing in MPEG SAOC

Kwangki KIM\(^{†}\), Student Member, Minsoo HAHN\(^†\), and Jinsul KIM\(^††\), Nonmembers

**SUMMARY**  MPEG spatial audio object coding (SAOC) is a new audio coding standard which efficiently represents various audio objects as a down-mix signal and spatial parameters. MPEG SAOC has a backward compatibility with existing playback systems for the down-mix signal. If a mastering signal is used for providing CD-like sound quality instead of the down-mix signal, an output signal decoded with the mastering signal may be easily degraded due to the difference between the down-mix and the mastering signals. To successfully use the mastering signal in MPEG SAOC, the difference between two signals should be eliminated. As a simple mastering signal processing, we propose a mastering signal processing using the mastering down-mix gain (MDG) which is similar to the arbitrary down-mix gain of MPEG Surround. Also, we propose an enhanced mastering signal processing using the MDG bias in order to reduce quantization errors of the MDG. Experimental results show that the proposed schemes can improve sound quality of the output signal decoded with the mastering signal. Especially, the enhanced method shows better performance than the simple method in the aspects of the quantization errors and the sound quality.

**key words:** MPEG SAOC, down-mix signal, mastering signal, mastering down-mix gain

1. Introduction

Recently, an interactive audio service has been introduced to satisfy consumers’ demands for an alternative and enhanced audio service[1], [2]. The interactive audio service delivers various audio objects such as bass, guitar, and vocal to users, so people can control specific audio objects and create their own audio sound scenes. This interactive audio service, however, has a constraint that a high bit-rate should be guaranteed because the various audio objects need to be separately coded. As an approach to solve the bit-rate problem in the interactive audio service, spatial audio object coding (SAOC) schemes have been introduced and developed[3], [4]. Especially, the MPEG Audio subgroup has been developing MPEG SAOC as a standard of the SAOC[5]–[8]. MPEG SAOC represents various audio objects as a down-mix signal and spatial parameters.

MPEG SAOC has a backward compatibility in which the down-mix signal is directly played by existing playback systems if there is no MPEG SAOC decoder or the users want to listen to the down-mix signal. In that sense, a mastering signal can be used as the down-mix signal because the mastering signal is a modified version of the down-mix signal and it has CD-like sound quality[9]. Although the mastering signal can provide the users with the CD-like sound quality in the aspect of the backward compatibility of MPEG SAOC, it may result in the performance degradation in the aspect of the MPEG SAOC decoding due to difference between the down-mix and the mastering signals. In this paper, a mastering signal processing is described as a method to successfully use the mastering signal in MPEG SAOC and an enhanced mastering signal processing is also presented to improve the performance of the mastering signal processing.

The rest of this paper is organized as follows. Section 2 summarizes MPEG SAOC briefly. In Sect. 3, the mastering signal and its relationship with MPEG SAOC are presented. In Sect. 4 and 5, the proposed simple mastering signal processing to support the use of the mastering signal in MPEG SAOC and its enhanced method are presented, respectively. Experimental results and the conclusion are given in Sect. 6 and 7.

2. Overview of MPEG SAOC

MPEG SAOC consists of the encoding, the decoding, and the transcoding part as shown in Fig. 1. In the encoding part, the input audio objects are represented as the down-mix signal and the spatial parameters. MPEG SAOC mainly uses object level difference (OLD) and inter-object correlation (IOC) as the spatial parameters. The OLD and the IOC are defined as the level difference and the amount of correlation or coherence among the input audio objects, respectively.

In the decoding part, the audio objects are reconstructed using the transmitted down-mix signal and the spatial parameters. Then, the desired output signal is generated by rendering the recovered objects according to the rendering matrix for the mono or the stereo output.

Since MPEG Surround is a well-defined multi-channel audio coding standard including many kinds of audio coding schemes for high sound quality, high coding efficiency, and backward compatibility, MPEG SAOC simply uses MPEG Surround[10] to generate the desired multi-channel output signals. In order to use MPEG Surround, the down-mix signal and the bit-stream for MPEG Surround decoding are necessary. The transcoding part of MPEG SAOC is originated for the use of MPEG Surround decoding. In the transcoding part, the MPEG SAOC bit-stream is transcoded into the MPEG Surround bit-stream according to the rendering matrix that indicates location and mixing information...
of audio objects at the playback setup or the desired output signals. Moreover, the down-mix signal of MPEG SAOC is pre-processed for optimum rendering performance. In other words, the down-mix signal and the bit-stream of MPEG SAOC are transformed into the pre-processed down-mix signal and the bit-stream for the MPEG Surround decoding according to the rendering matrix through the transcoding process of MPEG SAOC [4], [6].

Additionally, MPEG SAOC has the backward compatibility with existing playback systems for the down-mix signal. In other words, if there is no MPEG SAOC decoder/transcoder or the users want to listen to the down-mix signal, the transmitted down-mix signal can be simply played using the mono or stereo playback systems.

3. Mastering Signal and MPEG SAOC

Generally, a piece of music is sequentially produced through the processes such as the creating musical idea, the recording, the mixing, and the mastering as shown in Fig. 2. In other words, various audio objects such as bass, guitar, piano, vocal, and so on are recorded according to a composer’s intention, and then they are mixed into the stereo signal and mastered for the distribution as an audio CD. Especially, the mastering process is to make adjustments to the overall sound quality of the stereo mixing signal through the following mastering steps.

- **Edit:** Edit the stereo mixing signal to remove any unwanted sound from the beginning and end of the mixing signal.
- **Equalize:** Make any final tonal adjustments to the stereo mixing signal using equalizers.
- **Compress:** Make final adjustments to the overall level of the file using compression.
- **Distribute:** Save the final version as an MP3 file or burn to a recordable CD.

When MPEG SAOC is compared with the music production process, it can be concluded that MPEG SAOC can support all the music production process except the mastering process. It is because the input audio objects and the down-mix signal of MPEG SAOC are conceptually identical to the recorded audio objects and the mixing signal in the music production process. To cover the mastering process in MPEG SAOC, we propose that the mastering signal should be used as the down-mix signal.

As previously mentioned, MPEG SAOC has the backward compatibility with the existing playback systems for the down-mix signal. By using the mastering signal instead of the down-mix signal, MPEG SAOC can provide the users with the CD-like sound quality of the mastering signal. However, since the mastering signal is different from the down-mix signal as shown in Fig. 3, the performance of MPEG SAOC may be easily degraded in the aspects of the decoding and the transcoding of MPEG SAOC.
In order to successfully use the mastering signal for the down-mix signal without the performance degradation of MPEG SAOC, there should be a process to compensate the difference between the down-mix and the mastering signals. The compensation process can be performed using new parameters that describe how to modify the mastering signal for the decoding and the transcoding of MPEG SAOC. As a simple method to use the mastering signal in MPEG SAOC, the artistic down-mix concept of MPEG Surround can be applied [10]–[12].

4. Simple Mastering Signal Processing in MPEG SAOC and Its Problem

Similar to the arbitrary down-mix gain (ADG) in MPEG Surround, the mastering down-mix gain (MDG) is proposed as a new parameter having about 1.5 kb/s data rate and supporting the use of the mastering signal in MPEG SAOC. The MDG is defined as the power ratio between the down-mix signal generated by an MPEG SAOC encoder and the mastering signal created by a mastering engineer. The MDG is used to reconstruct a proper down-mix signal for the decoding and the transcoding processes by modifying the mastering signal.

Figure 4 shows the modified MPEG SAOC structure for the use of the mastering signal. Here, the modified MPEG SAOC transcoder structure is skipped. As shown in Fig. 4, in the encoder side, the MDG is additionally extracted through the down-mix analysis and it is formatted with other spatial parameters for the transmission. In the decoder side, the mastering signal is directly played or the SAOC decoding and transcoding are performed after the down-mix compensation using the MDG.

The MDG in partition b, denoted by $MDG_b$, is estimated as

$$MDG_b = 10 \log_{10} \frac{P_{m,b}}{P_{d,b}}, 1 \leq b \leq K,$$

(1)

where $P_{m,b}$ and $P_{d,b}$ are the estimated power of the mastering and down-mix signals, respectively, while $b$ is a parameter band index which indicates the spectral region containing several spectral components and $K$ is the number of parameter band [3]. Using $MDG_b$ and $P_{m,b}$, the power of down-mix signal can be obtained as

$$P_{d,b} = P_{m,b} \times \frac{1}{10^{MDG_b/10}} \text{ for } 1 \leq b \leq K.$$

(2)

The MDG is quantized for the transmission using the channel level difference (CLD) quantization scheme with non-uniform bidirectional 31 quantization levels of MPEG Surround as shown in Fig. 5. The CLD quantization scheme has fine resolutions in the region of small CLD values while coarse resolutions in the region of large CLD values, so it is suitable for the quantization of Gaussian distribution parameters with zero mean [10], [13]. Since the mastering signal can be considered as the modified version of the down-mix signal which is conceptually identical to the mixing signal in the music production process, it is natural to assume that the MDG has Gaussian-like distribution with near zero mean and it can be efficiently quantized using the CLD quantization table. However, though the MDG has Gaussian-like distribution, its mean is nonzero due to the effect of down-mix gains. In MPEG SAOC, a clipping problem of the down-mix signal causing sound quality degradation is not avoidable if all input objects are summed together without controlling each object gain [14]. To prevent the clipping, MPEG SAOC utilizes the down-mix gains smaller than 1 and these small gains lead to the power reduction of the down-mix signal. Thus, when the MDG is estimated, the MDG shows Gaussian distribution with nonzero mean. Namely, the center of the MDG distribution locates at the position far from 0 dB. Consequently, the MDG has large quantization errors.

5. Enhanced Mastering Signal Processing

To solve the MDG quantization problem, the MDG modification using the down-mix gains is proposed. From the fact that the down-mix gains result in the shift of the MDG distribution, it is devised that the center of the MDG distribution is moved back into the near zero position using the down-mix gains. The shift of the MDG distribution caused by the down-mix gains is defined as the MDG bias and it can be approximately calculated using MPEG SAOC parameters such as the down-mix gain (DMG) and the down-mix channel level difference (DCLD) which are estimated from the down-mix gains and transmitted to the decoder [6].

For the description of the MDG bias calculation, it is
assumed that there are the stereo down-mix signal, the mastering signal, and the down-mix matrix for obtaining the down-mix signal given as shown in (3).

\[
M_{\text{mix}} = \begin{bmatrix}
G_{l1} & G_{l2} & G_{l3} & \cdots & G_{lN} \\
G_{r1} & G_{r2} & G_{r3} & \cdots & G_{rN}
\end{bmatrix},
\]

(3)

where \(N\) is the number of objects and the subscript \(l\) and \(r\) denote the left and the right channel, respectively. The elements in the top and the bottom rows are the down-mix gains of the input objects for the left and the right channel of the down-mix signal, respectively. In the MPEG SAOC encoder, this down-mix matrix is parameterized as the DMG and the DCLD using (4).

\[
\begin{align*}
\text{DMG}_i &= 10 \log_{10} \left( \frac{G_{l,i}^2 + G_{r,i}^2}{G_{l,i} G_{r,i}} \right) \\
\text{DCLD}_i &= 20 \log_{10} \left( \frac{G_{l,i}}{G_{r,i}} \right) 
\end{align*}
\]

for \(1 \leq i \leq N\).

(4)

Using the DMG and the DCLD, all down-mix gains in (3) can be recalculated as

\[
\begin{align*}
\tilde{G}_{l,i} &= \sqrt{10^{\frac{\text{DCLD}_l}{10}} \times 10^{\frac{\text{DMG}_l}{20}}} \\
\tilde{G}_{r,i} &= \sqrt{10^{\frac{\text{DCLD}_r}{10}} \times 10^{\frac{\text{DMG}_r}{20}}}
\end{align*}
\]

for \(1 \leq i \leq N\),

(5)

where \(\tilde{G}_{l,i}\) and \(\tilde{G}_{r,i}\) are the recovered down-mix gains. In (5), the quantized DMG and the DCLD are used to avoid the mismatch between MDG biases calculated in the encoder and the decoder. The MDG bias can be obtained as (6) using (5).

\[
\begin{align*}
\text{MDG}_{l,i,\text{bias}} &= -10 \log_{10} \left( \frac{1}{N} \sum_{i=1}^{N} G_{l,i}^2 \right) \\
\text{MDG}_{r,i,\text{bias}} &= -10 \log_{10} \left( \frac{1}{N} \sum_{i=1}^{N} G_{r,i}^2 \right)
\end{align*}
\]

(6)

where \(\text{MDG}_{l,i,\text{bias}}\) and \(\text{MDG}_{r,i,\text{bias}}\) are the left and the right channel MDG biases, respectively. In (6), it is assumed that all objects have same unit power without respect to the characteristics of each object. It is because the effect of the down-mix gains on the power of the down-mix signal can be simply expected by the down-mix gains themselves. Based on this assumption, the reduced power of the down-mix signal by the down-mix gain can be approximately calculated as the squared mean of the down-mix gain. Finally, the modified MDG in partition \(b\) is calculated as

\[
\begin{align*}
m\text{MDG}_{l,b} &= \text{MDG}_{l,b} - \text{MDG}_{l,\text{bias}} \\
m\text{MDG}_{r,b} &= \text{MDG}_{r,b} - \text{MDG}_{r,\text{bias}}
\end{align*}
\]

for \(1 \leq b \leq K\),

(7)

where \(\text{MDG}_{l,b}\) and \(\text{MDG}_{r,b}\) are the original MDGs while \(m\text{MDG}_{l,b}\) and \(m\text{MDG}_{r,b}\) are the modified MDGs. These modified MDGs are quantized and transmitted to the decoder. Figure 6 shows the MDG distribution according to the down-mix gain and the modified MDG distribution by the MDG bias.

In the decoder side, the MDG bias is firstly recalculated using (5) and (6). The modified MDG is compensated with the calculated MDG bias as (8) and then used to modify the mastering signal using (2).

\[
\begin{align*}
\text{MDG}_{l,b} &= D\{ Q(m\text{MDG}_{l,b}) \} + \text{MDG}_{l,\text{bias}} \\
\text{MDG}_{r,b} &= D\{ Q(m\text{MDG}_{r,b}) \} + \text{MDG}_{r,\text{bias}}
\end{align*}
\]

for \(1 \leq b \leq K\),

(8)

where \(\text{MDG}_{l,b}\) and \(\text{MDG}_{r,b}\) are the recovered left and right channel MDGs. \(Q(\ )\) and \(D\{ \}\) denote the quantization and the dequantization processes. Figure 7 shows the
whole procedure of the enhanced mastering signal processing. As indicated in Fig. 7, the MDG bias is calculated only once during the encoding and the decoding process and the MDG modification is performed for every processing frame. Moreover, since the MDG modification only needs a few addition operations, the increase of the complexity by modifying the MDG is negligible.

6. Experimental Results

To confirm the effectiveness of the simple mastering signal processing and its enhanced method, two comparison tests on the aspects of the quantization errors and the subjective audio quality were performed. For all tests, 5 popular Korean songs were used and listed in Table 1. Each item consisted of one mastering signal and 4 to 6 audio object signals such as vocal, bass, guitar, drum, piano, etc. All test materials are available at [15].

Table 2 shows the quantization errors of the simple and the enhanced methods according to the variation of the down-mix gains. As shown in Table 2, the modified MDG has smaller quantization errors and its quantization errors are almost constant regardless of the variation of the down-mix gains. Here, we did not use the down-mix gains larger than 0.5 because the clipping of the down-mix signal occurred.

As a subjective listening test, the MUSHRA test was performed [16]. Six systems were used for the test and they are listed in Table 3. The down-mix signal was generated with 0.1 to 0.5 gains. Eight experienced listeners evaluated the decoded audio quality of the test items in each trial. As shown in Fig. 8, the subjective listening test results show that the sound quality of the output signals decoded with the mastering signals compensated by the MDG and the modified MDG are significantly better than that of the output signal generated with the original mastering signal and statistically similar to that of the output signal generated with the down-mix signal. Especially, the modified MDG shows slightly better sound quality compared to the MDG though the confidence intervals of two signals are almost overlapped. For the clarification, the differential score between the MDG and the modified MDG is given in Fig. 9. The modified MDG has the higher absolute score than the

| Table 1 | Test materials. |
|---|---|
| **Number** | **Material** | **List of object** |
| 1 | Hajimian | Guitar, bass, keyboard, rhythm, chorus, vocal |
| 2 | Braves | Guitar, bass, keyboard, rhythm, chorus, vocal |
| 3 | Snow | Guitar, bass, strings, rhythm, chorus, vocal |
| 4 | LaLaLa | Strings, bass, drum, vocal |
| 5 | SulpunDajim | Guitar, bass, piano&brass, rhythm, chorus, vocal |

| Table 2 | MDG quantization errors. |
|---|---|
| **Gain** | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | **Average** |
| MDG distortion (dB) | 1.11 | 1.17 | 1.27 | 1.47 | 1.68 | 1.34 |

| Modified MDG distortion (dB) | 1.05 | 1.06 | 1.05 | 1.09 | 1.10 | 1.07 |

| Table 3 | Six systems under test. |
|---|---|
| **Classification** | **Description** |
| Hidden reference | Reference signal generated with the original audio objects |
| Anchor | 3.5 kHz band-limited signal |
| Down-mix | Decoded output signal with the down-mix signal |
| Mastering | Decoded output signal with the mastering signal |
| MDG | Decoded output signal with the mastering signal compensated by the MDG |
| Modified MDG | Decoded output signal with the mastering signal compensated by the modified MDG |

Fig. 8 Subjective listening test results.

Fig. 9 Differential score between MDG and modified MDG.
MDG for all test items as shown in Fig. 9, although the differential score is not so high.

7. Conclusion

This paper presents mastering signal processing to support the use of the mastering signal in MPEG SAOC and its enhanced method with robustness to quantization errors. The simple mastering signal processing is designed to eliminate the difference between the down-mix and the mastering signals using the MDG. The MDG is very similar to the ADG of MPEG Surround, but it has the relatively large quantization errors using the MDG. The MDG is very similar to the ADG for all test items as shown in Fig. 9, although the differential score is not so high.

Although the MDG and the modified MDG show good performance, there still exists the sound quality degradation due to the power fluctuation problem caused by the sub-band domain processing. Therefore, it is needed to solve this problem for the improvement of sound quality. As one of possible solutions, the limitation to the dynamic range of the MDG in the low frequency region can be considered.

Acknowledgements

This work was supported by Chonnam National University and Brain Korea 21 Project, BK Electronics and Communications Technology Division, KAIST in 2012.

References

[1] D. Jang, T. Lee, Y. Lee, and J. Yoo, “A personalized preset-based audio system for interactive service,” 121st AES Convention, San Francisco 2006, Preprint 6904.
[2] ISO/IEC 23000-12, Information Technology–Multimedia application format (MPEG-A) – Part 12: Interactive music application format, 2010.
[3] C. Faller and R. Baumgarte, “Binaural cue coding-Part II: Schemes and application,” IEEE Trans. Speech Audio Process., vol.11, no.6, pp.520–531, Nov. 2003.
[4] J. Herre and S. Disch, “New concepts in parametric coding of spatial audio: From SAC to SAOC,” 2007 International Conference on Multimedia and Expo, pp.1894–1897, July 2007.
[5] ISO/IEC JTC1/SC29/WG11 (MPEG), Call for Proposals on Spatial Audio Object Coding, Document N8853, Jan. 2007.
[6] ISO/IEC 23003-2, Information Technology—MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC), 2010.
[7] J. Breebaart, J. Engdegard, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, J. Koppens, W. Oomen, B. Resch, E. Schuijers, and L. Terentiev, “Spatial audio object coding (SAOC) – The upcoming MPEG standard on parametric object based audio coding,” 124th AES Convention, Amsterdam 2008, Preprint 7377.
[8] K. Kim, J. Seo, S. Beack, K. Kang, and M. Hahn, “Spatial audio object coding with two-step coding structure for interactive audio service,” IEEE Trans. Multimedia, vol.13, no.6, pp.1208–1216, Dec. 2011.
[9] Berkleeshares.com, “Desktop music production for mac. lesson 1: Music production process,” Free music lessons from Berklee College of Music.
[10] ISO/IEC 23003-1, Information Technology—MPEG audio technologies – Part 1: MPEG Surround, 2007.
[11] ISO/IEC JTC1/SC29/WG11 (MPEG), Verification of MPEG surround artistic stereo downmix and next steps, Document M12566, Nice, Oct. 2005.
[12] ISO/IEC JTC1/SC29/WG11 (MPEG), MPEG Surround improved performance of the artistic downmix mode, Document M12899, Bangkok, Jan. 2006.
[13] K. Kim, S. Beack, J. Seo, D. Jang, and M. Hahn, “Improved channel level quantization for spatial audio coding,” ETRI J., vol.29, no.1, pp.99–102, Feb. 2007.
[14] H. Pang, “Clipping prevention scheme for MPEG surround,” ETRI J., vol.30, no.4, pp.606–608, Aug. 2008.
[15] [Online]. Available: ftp://143.248.158.110 (Contact us for ID and password).
[16] ITU-R Recommendation, Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA), ITU, BS. 1543-1, Geneva, 2001.

Kwangki Kim received the B.S. degree in electronic engineering from Korea Aviation University, Koyang, South Korea, in 2002, the M.S. degree in electronic engineering from Information and Communications University, Daejeon, South Korea, in 2004, and the Ph.D. degree in department of Information and Communications Engineering at Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011. He is currently with electrical engineering at KAIST. His research interests include multi-channel audio coding, multi-object audio coding and their applications.

Minsoo Hahn received the B.S. and the M.S. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 1979 and 1981, respectively, and the Ph.D. degree in electrical and electronics engineering from University of Florida, Florida, USA, in 1989. From 1990 to 1997, and he was with Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. In 1998, he has been a faculty member of the School of Engineering, Information and Communications University. Currently, he is a full professor in electrical engineering at Korea Advanced Institute of Science and Technology (KAIST) and a Director in Digital Media Laboratory, KAIST. His research interests include speech and audio coding, speech synthesis, noise reduction and VoIP.
Jinsul Kim received the B.S. degree in computer science from University of Utah, Salt Lake City, Utah, USA, in 2001, and the M.S. and Ph.D. degrees in digital media engineering, department of information and communications from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2005 and 2008. He worked as a researcher in IPTV Infrastructure Technology Research Laboratory, Broadcasting/Telecommunications Convergence Research Division, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea from 2005 to 2008. He worked as a professor in Korea Nazarene University, Chon-an, Korea from 2009 to 2011. Currently, he is a professor in Chonnam National University, Gwangju, Korea. He has been invited reviewer for IEEE Trans. Multimedia since 2008. He has been invited for TPC (Technical Program Committee), IWITMA2009/2010, and PC (Program Chair), ICCCT2011 His research interests include QoS/QoE, Measurement/Management, IPTV, Mobile IPTV, Smart TV, Multimedia Communication and Digital Media Arts.