International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.2 (2014), pp.253-258
http://dx.doi.org/10.14257/ijmue.2014.9.2.25

Sound Scene Control of Multi-Channel Audio Signals for Realistic Audio Service in Wired/Wireless Network

Kwangki Kim

Department of Digital Contents, Korea Nazarene University
k2kim@kornu.ac.kr

Abstract

In this paper, we proposed the sound scene control (SSC) of the multi-channel audio signals in MPEG Surround to enhance the MPEG Surround with respect to the functionality. To add a new functionality to the MPEG Surround, we proposed the multi-channel sound scene control by modification of the spatial parameters such as the channel level difference and the inter-channel correlation. Even though the panning law is directly applied to the spatial parameters, the output channel signals are successfully panned by given panning angle. The quality and the localization ability of the proposed SSC are confirmed by the subjective listening test.

Keywords: *MPEG Surround, multi-channel audio signal, spatial parameter, sound scene control*

1. Introduction

MPEG Surround is a technology that multi-channel audio signals are represented as a down-mix signal and spatial parameters [1-3]. As the MPEG Surround only needs bit-rate of the down-mix signal and the additional side information for the transmission, the multi-channel audio signals can be delivered to users through wired/wireless network system by the MPEG Surround. Therefore, users can enjoy more realistic audio sound through digital audio broadcasting, digital multimedia broadcasting and so on in wired/wireless network environment. The MPEG Surround uses channel level difference (CLD) and inter-channel correlation (ICC) as spatial parameters. The CLD is a main parameter in the MPEG Surround and it determines the spectral power of the reconstructed multi-channel audio signals [4]. Whereas, the ICC is an ancillary parameter in the MPEG Surround and it reflects the spatial diffuseness of the recovered multi-channel audio signals. As the multi-channel audio sound are compressed and recovered using the down-mix signal and the spatial parameters, the performance of the MPEG Surround in the aspects of the coding efficiency and the sound quality is determined by the spatial parameters.

Apart from the coding efficiency and the sound quality determined in the MPEG Surround by the spatial parameters, there are many merits able to create valuable functionality through the usage of spatial parameters. Since the spatial image of the multi-channel audio signals can be preserved by the spatial parameters, we can control the sound scene of the multi-channel audio signals by the modification of the spatial parameters. We call this sound scene control (SSC) based on the spatial parameters. One of possible applications of the sound scene control is to combine sound scene controller with multi-view video. In the case of multi-view broadcasting, which will be advent in near coming days, it is expected that multi-channel sound scene control can

ISSN: 1975-0080 IJMUE

provide interactive audio playback systems and realistic audio sound by synchronizing sound scene with moving video scene. In this paper, we present the SSC by the modification of the spatial parameters with low complexity.

2. Basic Concept of Sound Scene Control

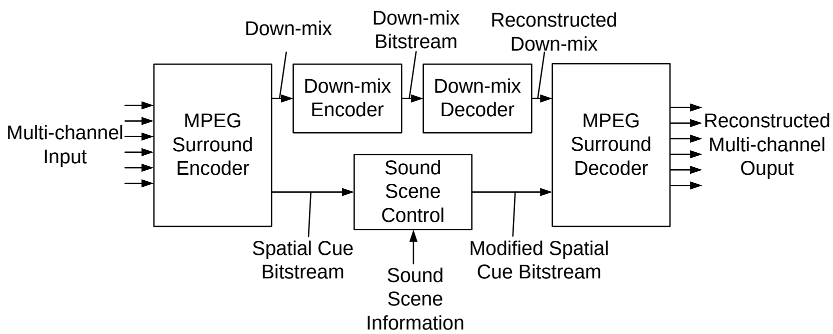

Figure 1. MPEG Surround with SSC

The SSC is a new tool to reproduce a new sound scene of the multi-channel audio signals according to the global panning position which is freely inputted by a user or other system. By given panning angle, denoted by θ_{pan}, the multi-channel audio signals are rotated with a degree of θ_{pan}. As previously mentioned, the SSC can be simply realized by modification of the spatial cues. Namely, the SCC modifies spatial cues according to inputted sound scene information and then generates the modified spatial cues. These modified spatial cues are delivered to the MPEG Surround decoder and the multi-channel audio signals with controlled sound scene are reconstructed through decoding process. Figure 1 shows the structure of MPEG Surround with the proposed SSC.

3. Procedure of Sound Scene Control

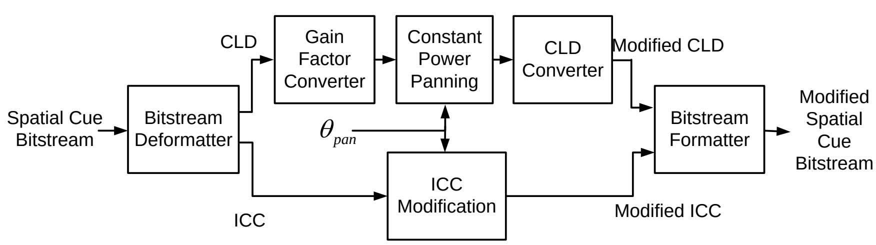

Figure 2. Procedure of SCC

The procedure of the SSC in the MPEG Surround is shown in Figure 2. At first, the spatial parameters such as the CLD and the ICC are parsed from the transmitted spatial parameter bit-stream. And then they are modified according to θ_{pan} which is sound scene information. Here, the CLD and the ICC are separately controlled and it is assumed that θ_{pan} is fed to each modification module. These modified spatial parameters are formatted again and finally the modified spatial parameter bit-stream is generated. The modification methods of spatial parameters are explained in the following sections.

3.1 CLD modification

The CLD plays a pivotal role in the SCC because it is used for the estimation of power gain of each channel corresponding to the desired sound scene. To modify the CLD, it is processed by gain factor converter, constant power panning (CPP), and CLD converter, sequentially. In the gain factor converter, the CLD is converted to each channel level gain per each sub-band. The gain factors are simply calculated from the CLD as following formula.

$$G_b^i = \frac{1}{\sqrt{1+10^{CLD_b/10}}} \tag{1}$$

$$G_b^{i+1} = G_b^i \cdot 10^{CLD_b/20} \tag{2}$$

where the G_b^i is the gain factor, the superscription index i is channel index, and subscription index b is sub-band index. It is known that one CLD per sub-band can provide two channel's power gain. This gain converter is applied to all CLDs and each channel level gain can be easily obtained by multiplying all gain factors related to each channel.

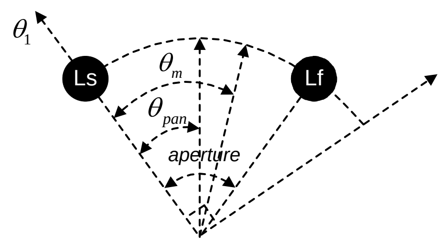

Figure 3. CPP law between two channels

In CPP module, the CPP law is applied to manipulate the position of each channel according to desired sound scene [5], [6]. Let's assume that if the channel gain G_b^i is desired to be positioned at θ_{pan} which is located at between left front (Lf) and left surround (Ls) as shown in Figure 3, the G_b^i is projected to Lf and Ls channels as follows:

$$\theta_m = \frac{(\theta_{pan} - \theta_1)}{(aperture - \theta_1)} \times \frac{\pi}{2} \tag{3}$$

$$G_{b,new}^{Lf} = G_b^{Lf} + \cos(\theta_m) \cdot G_b^i$$
$$G_{b,new}^{Ls} = G_b^{Ls} + \sin(\theta_m) \cdot G_b^i \tag{4}$$

where θ_m is the normalized angle limited to 90 degree and aperture is the angle between two channels. As the same manner, any other channel gain can be flexibly handled to form the desired sound scene. After the CPP processing, the modified CLDs are newly estimated using all new channel gains in CLD converter. Here, the CLD converter is exactly same as the CLD extractor of the MPEG Surround encoder. If the CLD is estimated between Lf and Ls channels, the modified CLD is calculated as follows.

$$CLD_{b,new}^{Lf,Ls} = 10\log_{10}\left(\frac{G_{b,new}^{Ls}}{G_{b,new}^{Lf}}\right)^2 \tag{5}$$

3.2 ICC modification

The ICC has an additional role in the SCC because it reflects the changed spatial diffuseness corresponding to desired sound scene. To perfectly modify the ICC, the ICC must be re-estimated according to the controlled sound scene. But, different from the CLD, the ICC cannot be re-estimated in parameter domain since the degree of correlation between the channels is only able to be estimated in signal domain.

Due to this problem, the ICC cannot be perfectly controlled and it could result in the degradation of overall sound quality after changing the sound scene. Despite of this restriction, two kinds of ICC parameter could be modified in the case of sound scene rotation.

$$ICC_{Ls,Lf} = (1-\eta)ICC_{Ls,Lf} + \eta ICC_{Rs,Rf} \tag{6}$$

$$ICC_{Rs,Rf} = (1-\eta)ICC_{Rs,Rf} + \eta ICC_{Ls,Lf} \tag{7}$$

where η is denoted as,

$$\eta = \begin{cases} \dfrac{\theta_{pan}}{\pi} \quad , & \theta_{pan} \le \pi \\ 1 - \dfrac{\theta_{pan} - \pi}{\pi}, & \theta_{pan} > \pi \end{cases} \tag{8}$$

These equations mean that left and right half plane ICC parameters are totally cross-changed if the degree of scene rotation is equal to the 180 degree. In the case that rotation angle is increased greater than 180 degree to 360 degree, the reverse cross-changed is occurred and finally equal to the original at 360 degree. This concept of modification is originated from common smoothing technique used in the MPEG Surround [1, 3].

4. Complexity of Sound Scene Control

Complexities between proposed method and direct panning of decoded multi-channel signals (called ‘Refpan’) are compared. The used panning method in parameter modification is also identically applied to produce ‘Refpan’ signal. The comparison of complexity is shown in Table 1. The addition operation is not considered and the number of multiplication in FFT operation is referred to [7]. From the comparison, the proposed one can save considerable computations compared with ‘Refpan’.

Table 1. Comparison of complexity between CLD modification based panning and reference panning. (Number of sub-band: 28, Frame size: 2048)

Classification	Proposed	Refpan
Multiplication	624	4096
Trigonometrical operation	equal	Equal
Logarithm	132	-
Multiplication in FFT	-	22528
division	806	140

5. Experiment

Table 2. Test items

Index	Material	Description
A	Applause	Ambience
B	ARL_applause	Ambience
C	Chostakovitch	Music (back: direct)
D	Crit	Pathological (back: ambience)
E	Fountain_music	Pathological
F	Glock	Pathological
G	Indie2	Movie sound
H	Jackson1	Music (back: ambience)
I	Pops	Music (back: direct)
J	Poulenc	Music (back: direct)
K	Rock_concert	Music (back: ambience)
L	Stomp	Movie sound

To validate the performance of the proposed SSC, the subjective listening test was performed. For listening tests, the twelve test items offered by MPEG Audio sub-group were used and are listed in Table 2. To simplify the test, the desired sound scene is only set to 90 degree since that position is considered as critical point where the sound quality could be degraded. A MUSHRA test was used as a test methodology and eight experienced subjects were participated in the test [8]. Three systems were used for the test and they are 'Hidden ref', 'RMP', and 'Proposed'. 'Hidden ref' are directly panned original multi-channel signals and 'RMP' are directly panned decoded signals in MPEG Surround. Whereas, 'Proposed' are decoded signals of MPEG Surround panned using SSC. Figure 4 shows the subjective listening test results. The proposed method is slightly degraded compared with RMP, but still preserve statistically same sound quality with the 95% confidence interval. In other words, the degradation by the proposed SSC is negligible.

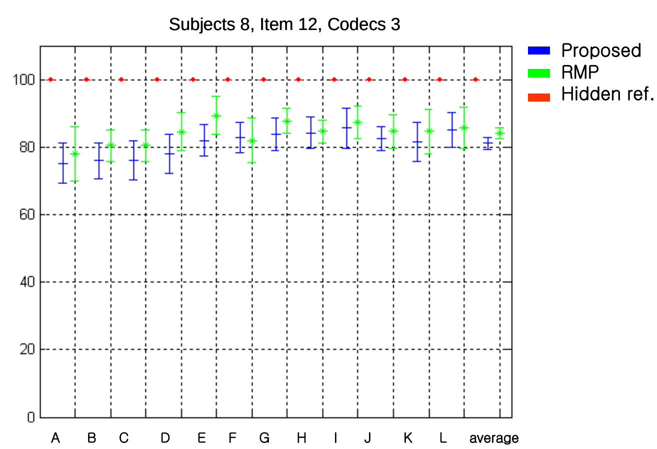

Figure 4. Subjective listening test results

6. Conclusion

In this paper, we proposed the SSC of the multi-channel audio signals to enhance the MPEG Surround with respect to the functionality. To add a new functionality to the MPEG Surround, we proposed the multi-channel sound scene control by modification of the spatial parameters such as the CLD and the ICC. Even though the panning law is directly applied to the spatial parameters, the output channel signals are successfully panned by given panning angle. The quality of the proposed SSC is confirmed by the subjective listening test.

Acknowledgements

This study was funded by the research fund of Korea Nazarene University in 2013 (2013-0302)

References

[1] ISO/IEC 23003-1, Information Technology–MPEG AudioTechnologies – Part 1: MPEG Surround, **(2007)**.
[2] ISO/IEC JTC1/SC29/WG11 (MPEG), Call for Proposals on Spatial Audio Object Coding, Document N8853, **(2007)** January.
[3] J. Herre, H. Purnhagen, J. Breebard, C. Faller, S. Disch, K. Kjorling, E. Schuijers, J. Hilpert and F. Myburg, "The Reference Model Architechture for MPEG Spatial Audio Coding", 118th AES Convention, Barcelona **(2005)**, Preprint 6447.
[4] K. Kim, S. Beack, J. Seo, D. Jang and M. Hahn, "Improved Channel Level Difference Quantization for Spatial Audio Coding," ETRI Journal, vol. 29, no. 1, **(2007)** February, pp. 99-102.
[5] V. Pulkki, "Virtual sound source positioning using vector base amplityde panning", Jl. Audio Eng. Soc., vol. 45, **(1997)** June, pp. 456-466.
[6] M. A. Gerzon, "Panpot laws for multispeaker stereo", 92nd Convention of the AES, J Audio Eng. Soc., Preprint 3309, **(1992)** May.
[7] S. Haykin, "Adaptive Filter Theory", 4th Ed. Prentice-Hall.
[8] ITU-R Recommendation, Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA), ITU, BS. 1543-1, Geneva, **(2001)**.

Author

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