



A Study on Sound Scene Control of Binaural Sound in Multi-channel Audio Coding

Kwangki Kim*

Department of Digital Contents, Korea Nazarene University

ABSTRACT

Recently, realistic visual services such as 3DTV, UHDTV, 3D cinema, 3D gaming, augmented reality (AR), virtual reality (VR) and so on are introduced, and realistic audio services are getting more important to make the realistic visual services more realistic. Generally, the realistic audio services have a limitation that users should have multi-channel speaker layout for generating the realistic audio sound, so the binaural sound generation method that provides the users the realistic sound through a stereo headphone has been introduced. But, since the binaural sound has a constant sound scene, it can not reflect the users' movement in the AR or the VR service. In this paper, we proposed a sound scene control scheme of binaural sound in the multi-channel audio coding for generating a more realistic audio sound in the VR or AR service. In the proposed method, multi-channel gain factors and a stereo down-mix signal are modified based on the new multi-channel layout to be arranged according to a user's head movement. The experimental results show that the proposed sound scene control scheme of the binaural sound in the multi-channel audio coding can successfully generate the binaural sound with the controlled sound scene while maintaining the good sound quality.

© 2017 KKITS All rights reserved

KEYWORDS : Realistic visual service, Realistic audio service, Virtual reality, Binaural sound, Sound scene control

ARTICLE INFO: Received 23 January 2017, Revised 5 February 2017, Accepted 10 February 2017.

*Corresponding author is with the Department of Digital Contents, Korea Nazarene University, 48 Wolbong-ro Seobuk-gu Cheonan-si Chungcheongnam-do, 31172,

S.KOREA.

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

1. Introduction

Nowadays, there are many realistic visual services such as 3DTV, UHDTV, 3D cinema, 3D gaming, augmented reality (AR), virtual reality (VR) and so on [1, 2]. To make these services more realistic, realistic audio services should be adopted and are getting more important in the realistic visual services. Generally, the realistic audio services provide users with the realistic audio sound to be generated by more than 5.1 multi-channel audio signals. So, the users must use the multi-channel speaker layout for enjoying the realistic audio sound. However, since most of users do not have the multi-channel speak layout and they can experience the realistic audio sound only in the special places such as a music hall and a theater, the realistic audio service cannot be widely supplied. As one of solutions to the limited realistic audio service, a binaural sound can be utilized. The binaural sound is a stereo signal that is generated by convolving the multi-channel audio signal with HRTF (head related transfer function) coefficients [3, 4]. Since the HRTF is a response that characterizes how a sound in free space reaches an ear, the binaural sound has a multi-channel audio effect and the users can enjoy the realistic audio sound using a stereo headphone.

Meanwhile, although more than 5.1 channel audio signals can generate the realistic audio sound, they cannot be directly delivered to the users through the wired and wireless network system due to the high data rate. So, the multi-channel audio coding methods such as

binaural cue coding, MPEG Surround, and sound source location coefficient coding (SSLCC) are used to solve the bit-rate problem of the realistic audio service [5-11]. Generally, the multi-channel audio coding compresses the multi-channel audio signals as a stereo down-mix signal and additional side information. If the users have the multi-channel audio layout, they enjoy the realistic audio sound using the multi-channel audio signals recovered by the multi-channel audio coding. Otherwise, the users only listen to the stereo down-mix signal without any multi-channel audio effect. To overcome the drawback of the stereo down-mix signal without the realistic audio effect, the binaural decoding is adopted and the users enjoy the realistic audio sound from the binaural signal generated by the binaural decoding instead of the stereo down-mix signal. However, since the binaural sound is generated based on the fixed multi-channel layout and the HRTF coefficients, the audio sound in the VR service cannot follow the users' movement if the multi-channel audio coding with the binaural decoding is applied to the VR service. In other words, the binaural sound has a constant sound scene, so the audio sound in the VR service is not accompanied with the visual scene in the VR if the users move to anywhere in the VR. From this reason, the binaural sound should be modified for reflecting the users' position change. In this paper, we proposed the sound scene control scheme of the binaural sound in the multi-channel audio coding to allow that users' free movement in the VR service. In the proposed method, multi-channel gain factors and a

stereo down-mix signal are modified based on the new multi-channel layout to be arranged according to users' head movement and then the binaural sound is generated using the modified multi-channel gain factors and the stereo down-mix signal.

2. Binaural Decoding in SSLCC

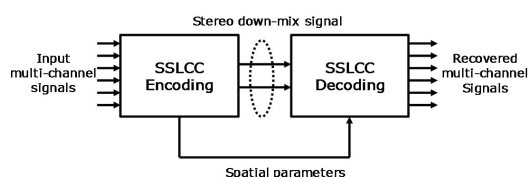


그림 1. SSLCC의 전체 구조
Figure 1. Overall structure of the SSLCC

The SSLCC is a technology that efficiently handles the multi-channel audio signals with a good performance in the aspects of bit-rate and sound quality. Especially, the SSLCC compresses the multi-channel audio signals with respect to a sound image in the free space. <Figure 1> shows an overall structure of the SSLCC. In the SSLCC encoder, the multi-channel audio signals are down-mixed into a stereo signal and spatial parameters are extracted for reconstructing the multi-channel audio signals. Here, a virtual source location information (VSLI) is used as the spatial parameters and it indicates the power information of each single channel of the multi-channel signals. Moreover, the VSLI is an angle of the spatial image in free space to be formed by the multi-channel audio signals. In the decoder side, the multi-channel audio signals are recovered by using the received stereo down-mix signal and the

spatial parameters. To generate the multi-channel audio signals, gain factors of the multi-channel signals are calculated from the VSLI and they are multiplied by the stereo down-mix signal in the frequency domain. The detailed explanation of the SSLCC encoding and decoding can be found in [6] and [7].

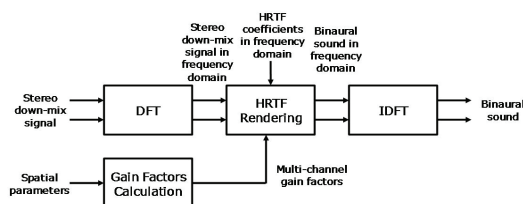


그림 2. SSLCC에서의 바이노럴 복호화 과정
Figure 2. Binaural decoding process in SSLCC

<Figure 2> shows a process of binaural decoding in the SSLCC. For the reduction of the computational loads of binaural decoding caused by the linear convolution, the binaural decoding is performed in the frequency domain. At first, the HRTF coefficients are pre-transformed in the frequency domain by the DFT and stored in the memory. Then, the gain factors of the multi-channel signals are calculated from the spatial parameters in the frequency domain. The estimated gain factors of the multi-channel signals are multiplied by the HRTF coefficients, and these results are multiplied by the down-mix signal for obtaining the binaural sound in the frequency domain. Finally, the binaural sound in the frequency domain is transformed into the time domain by the IDFT. The detailed information on the binaural decoding process in the SSLCC can be found in [3] and [4].

For more explanation, let's assume that there are the calculated gain factors of the multi-channel signals, $g_{1c}(k), g_{2c}(k), g_{lf}(k), g_{ls}(k), g_{rf}(k), g_{rs}(k)$, the HRTF coefficients, $H_C^L(k), H_C^R(k), H_{Lf}^L(k), H_{Lf}^R(k), H_{Ls}^L(k), H_{Ls}^R(k), H_{Rf}^L(k), H_{Rf}^R(k), H_{Rs}^L(k), H_{Rs}^R(k)$, and the down-mix signal, $X_L(k), X_R(k)$. Using these signals and coefficients, the binaural decoding is performed as

$$\begin{aligned} H_{LL}(k) &= g_{1c}(k) \cdot H_C^L(k) + g_{lf}(k) \cdot H_{Lf}^L(k) + g_{ls}(k) \cdot H_{Ls}^L(k) \\ H_{RL}(k) &= g_{2c}(k) \cdot H_C^L(k) + g_{rf}(k) \cdot H_{Rf}^L(k) + g_{rs}(k) \cdot H_{Rs}^L(k) \\ H_{LR}(k) &= g_{1c}(k) \cdot H_C^R(k) + g_{lf}(k) \cdot H_{Lf}^R(k) + g_{ls}(k) \cdot H_{Ls}^R(k) \\ H_{RR}(k) &= g_{2c}(k) \cdot H_C^R(k) + g_{rf}(k) \cdot H_{Rf}^R(k) + g_{rs}(k) \cdot H_{Rs}^R(k) \end{aligned} \quad (1)$$

where $H_{LL}(k)$ and $H_{LR}(k)$ are HRTF elements for left and right binaural output by center (C), left front (Lf), and left surround (Ls) channels, respectively, while $H_{RL}(k)$ and $H_{RR}(k)$ are HRTF elements for left and right binaural output by C, right front (Rf), and right surround (Rs) channels, respectively. Here, k is a frequency index. Finally, the binaural output signals in the frequency domain is obtained as

$$\begin{bmatrix} O_L(k) \\ O_R(k) \end{bmatrix} = \begin{bmatrix} H_{LL}(k) & H_{LR}(k) \\ H_{RL}(k) & H_{RR}(k) \end{bmatrix} \cdot \begin{bmatrix} X_L(k) \\ X_R(k) \end{bmatrix} \quad (2)$$

where $O_L(k)$ and $O_R(k)$ are the left and right binaural output signals in the frequency domain, respectively.

3. Proposed Sound Scene Control of Binaural Sound

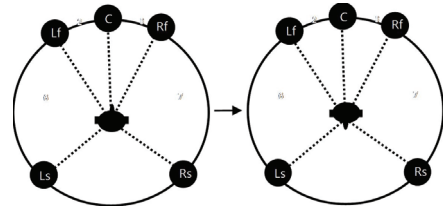


그림 3. 사용자의 위치(방위각) 변화 예
Figure 3. Example of user's head movement

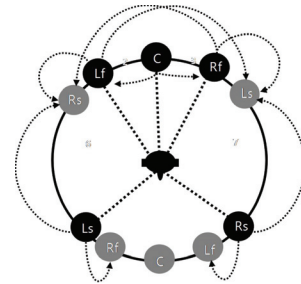


그림 4. 채널 매핑에 의해 생성된 다채널 신호 예
Figure 4. Example of new multi-channel signals generated by channel mapping

The sound scene control is a scheme that provide a user with a more realistic audio sound in the VR service. The audio sound in the VR service should be changed according to the user's head position, when they want to move anywhere in the VR. But, the binaural sound to be played with the stereo headphone only provide a constant sound scene, so it cannot be directly applied to the VR service that allows the user's head movement. Therefore, the binaural sound should be controlled for reflecting the user's head movement in the VR. <Figure 3> shows an example of the user's head movement in 5.1 channel layout. When the user's head moves to the opposite side - 180 degrees, Ls and Rs channels are perceived as the front channels while C, Lf, and Rf are noticed as the back channels.

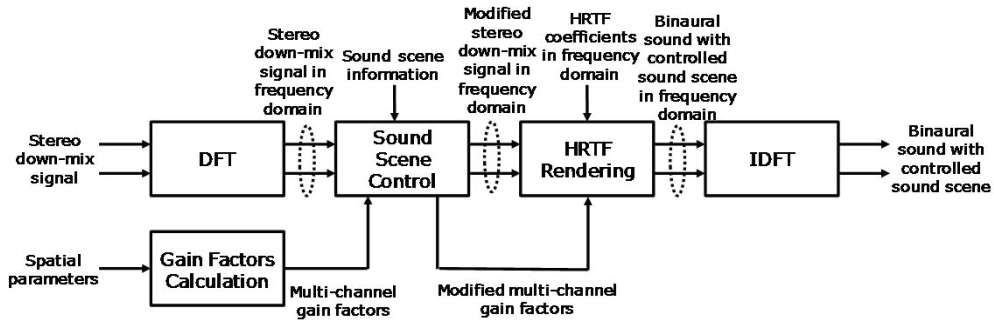


그림 5. 제안된 사운드스캔 조절 방법을 갖는 바이노럴 복호화 과정
 Figure 5. Proposed binaural decoding process with sound scene control

Therefore, to generate the binaural sound, all multi-channel signals should be mapped on a new multi-channel layout as shown in <Figure 4>. Then, the binaural sound can be generated by convolving the newly generated multi-channel signals with the HRTF coefficients.

<Figure 5> shows a proposed binaural decoding process with the sound scene control. In the previous binaural decoding process, the sound scene control module is inserted before the HRTF rendering. In the sound scene control module, the calculated gain factors of the multi-channel signals and the stereo down-mix signal are modified according to the sound scene information to be users' position. Then, the binaural sound with the controlled sound scene is generated with the modified down-mix signal and the multi-channel gain factors in the HRTF rendering module.

Angle information of the multi-channel layout is used for the sound scene control. <Figure 6> shows the angle information of the multi-channel layout and is the angle of the user's head movement. According to the user's head

movement, the angle of the multi-channel layout is changed as shown in <Figure 6>. Then, the original multi-channel audio signals are mapped onto the new multi-channel layout for generating the new multi-channel audio signals. The constant power panning (CPP) law is applied to generate the new multi-channel audio signals [12, 13].

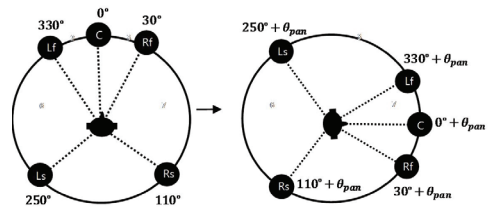


그림 6. 사용자 위치 변화에 따른 다채널 레이아웃 예
 Figure 6. Example of new multi-channel layout according to user's head movement

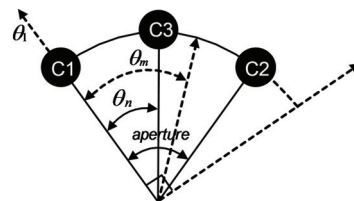


그림 7. Constant power panning 기법에 기반한 채널 매핑 예
 Figure 7. Example of channel mapping based on constant power panning law

<Figure 7> show an example of the channel mapping based on the CPP law. When a channel C3 lays in between channels C1 and C3 as the result of the user's head movement (θ_n), C3 is mapped onto both C1 and C2. Given the channel gain of C1($g_1(k)$), C2($g_2(k)$), and C3($g_3(k)$), the $G^3(k)$ is projected to C1 and C2 as follows:

$$\begin{aligned} \theta_m &= \frac{(\theta_n - \theta_1)}{(\text{aperture} - \theta_1)} \times \frac{\pi}{2}, \\ g_1^{new}(k) &= g_1(k) + \cos(\theta_m) \cdot g_3(k) \\ g_2^{new}(k) &= g_2(k) + \sin(\theta_m) \cdot g_3(k) \end{aligned} \quad (3)$$

where θ_m is the normalized angle less than 90 degrees and *aperture* is the angle between C1 and C2. In the same way, the new multi-channel gains can be calculated using the new multi-channel layout and the original multi-channel gains. When the angle of the user's head movement is 90 degree, the new multi-channel gains are calculated as follows:

$$\begin{aligned} g_{1c}^{new}(k) &= 0.707 \times \cos\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k) \\ g_{2c}^{new}(k) &= 0.707 \times \cos\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k) \\ g_{lf}^{new}(k) &= \sin\left(\frac{30-(-20)}{60-(-20)} \times \frac{\pi}{2}\right) \cdot g_{Rf}(k) + \sin\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot (g_{1c}(k) + g_{2c}(k)) \\ g_{Rf}^{new}(k) &= \sin\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k) \\ g_{Ls}^{new}(k) &= \cos\left(\frac{30-(-20)}{60-(-20)} \times \frac{\pi}{2}\right) \cdot g_{Rf}(k) + \cos\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot (g_{1c}(k) + g_{2c}(k)) \\ &\quad + \sin\left(\frac{340-330}{340-200} \times \frac{\pi}{2}\right) \cdot g_{lf}(k) + \sin\left(\frac{340-250}{340-200} \times \frac{\pi}{2}\right) \cdot g_{Ls}(k) \\ g_{Rs}^{new}(k) &= \cos\left(\frac{340-330}{340-200} \times \frac{\pi}{2}\right) \cdot g_{lf}(k) + \cos\left(\frac{340-250}{340-200} \times \frac{\pi}{2}\right) \cdot g_{Ls}(k) \end{aligned} \quad (4)$$

Meanwhile, the stereo down-mix signal should be also modified to reflect the user's head movement because the binaural sound is generated using the

multi-channel gain factors, the HRTF coefficients, and the stereo down-mix signal. When the multi-channel audio signal is mapped onto the new multi-channel layout, the stereo down-mix signal is newly generated using the mapped multi-channel gain factors. The new stereo down-mix signal is calculated as follows:

$$\begin{bmatrix} X_L^{new}(k) \\ X_R^{new}(k) \end{bmatrix} = \begin{bmatrix} g_{ll}(k) & g_{rl}(k) \\ g_{lr}(k) & g_{rr}(k) \end{bmatrix} \cdot \begin{bmatrix} X_L(k) \\ X_R(k) \end{bmatrix} \quad (5)$$

where $g_{ll}(k)$ and $g_{rl}(k)$ are the elements for the new left down-mix signal that come from the original left and right down-mix signals, respectively, while $g_{rr}(k)$ and $g_{lr}(k)$ are the elements for the new right down-mix signal that come from the original right and left down-mix signal. When the angle of the user's head movement is 90 degrees, four elements for the new stereo down-mix signal is calculated using the elements in (5) as follows:

$$\begin{aligned} g_{ll}(k) &= \sqrt{\left(\sin\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot g_{lc}(k)\right)^2 + \left(\cos\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot g_{lc}(k)\right)^2} \\ &\quad + \sqrt{\left(\sin\left(\frac{340-330}{340-200} \times \frac{\pi}{2}\right) \cdot g_{lf}(k)\right)^2 + \left(\sin\left(\frac{340-250}{340-200} \times \frac{\pi}{2}\right) \cdot g_{Ls}(k)\right)^2} \\ g_{rl}(k) &= \sqrt{\left(0.707 \times \cos\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k)\right)^2 + \left(\sin\left(\frac{30-(-20)}{60-(-20)} \times \frac{\pi}{2}\right) \cdot g_{Rf}(k)\right)^2} \\ &\quad + \sqrt{\left(\sin\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot g_{2c}(k)\right)^2 + \left(\cos\left(\frac{30-0}{60-0} \times \frac{\pi}{2}\right) \cdot g_{2c}(k)\right)^2} \\ g_{rr}(k) &= \sqrt{\left(0.707 \times \cos\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k)\right)^2 + \left(\sin\left(\frac{110-90}{120-90} \times \frac{\pi}{2}\right) \cdot g_{Rc}(k)\right)^2} \\ g_{lr}(k) &= \sqrt{\left(\cos\left(\frac{340-330}{340-200} \times \frac{\pi}{2}\right) \cdot g_{lf}(k)\right)^2 + \left(\cos\left(\frac{340-250}{340-200} \times \frac{\pi}{2}\right) \cdot g_{Ls}(k)\right)^2} \end{aligned} \quad (6)$$

Finally, (1) and (2) for generating the new binaural sound can be re-written as follows:

$$\begin{aligned}
 H_{LL}^{new}(k) &= g_{1C}^{new}(k) \cdot H_C^L(k) + g_{L_f}^{new}(k) \cdot H_{L_f}^L(k) + g_{L_s}^{new}(k) \cdot H_{L_s}^L(k) \\
 H_{RL}^{new}(k) &= g_{2C}^{new}(k) \cdot H_C^L(k) + g_{R_f}^{new}(k) \cdot H_{R_f}^L(k) + g_{R_s}^{new}(k) \cdot H_{R_s}^L(k) \\
 H_{LR}^{new}(k) &= g_{1C}^{new}(k) \cdot H_C^R(k) + g_{L_f}^{new}(k) \cdot H_{L_f}^R(k) + g_{L_s}^{new}(k) \cdot H_{L_s}^R(k) \\
 H_{RR}^{new}(k) &= g_{2C}^{new}(k) \cdot H_C^R(k) + g_{R_f}^{new}(k) \cdot H_{R_f}^R(k) + g_{R_s}^{new}(k) \cdot H_{R_s}^R(k)
 \end{aligned}
 \tag{7}$$

$$\begin{bmatrix} O_L^{new}(k) \\ O_R^{new}(k) \end{bmatrix} = \begin{bmatrix} H_{LL}^{new}(k) & H_{RL}^{new}(k) \\ H_{LR}^{new}(k) & H_{RR}^{new}(k) \end{bmatrix} \cdot \begin{bmatrix} X_L^{new}(k) \\ X_R^{new}(k) \end{bmatrix}
 \tag{8}$$

4. Experimental Results

To check the performance of the proposed binaural decoding with sound scene control, the subjective listening test was performed. For the test, three multi-channel audio contents were used and listed in Table 1 [14]. All test items have 44.1 kHz sampling rate and 16-bit resolution. A MUSHRA test was used as a test methodology and six experienced subjects were participated in the test [15]. Three systems were used and listed in Table 2. To simplify the test, the angle of the user’s head movement is only set to 90 degree.

<Figure 8> shows the subjective listening test results. For all test items, ‘DECODED’ shows a slightly better sound quality than ‘PROPOSED’ and has a slightly higher absolute score than ‘PROPOSED’. Nevertheless, since the confidence interval of ‘DECODED’ and ‘PROPOSED’ is overlapped, they have the statistically same sound quality. From the experimental results, it is confirmed that the proposed binaural decoding process with the sound scene control can successfully generate the binaural sound with the controlled sound scene while maintaining the good sound quality.

표 1. 실험 데이터

Table 1. Test materials

Material	Description
ARL_applause	Ambience
Chostakovitch	Music (back: direct)
Fountain_music	Pathological

표 2. 실험 대상

Table 2. System under test

Classification	Description
REF	Binaural sound generated with directly panned original multi-channel signals
DECODED	Binaural sound generated with directly panned decoded multi-channel signals in the SSLCC
PROPOSED	Binaural sound generated with the sound scene controlled multi-channel signals using SSC module

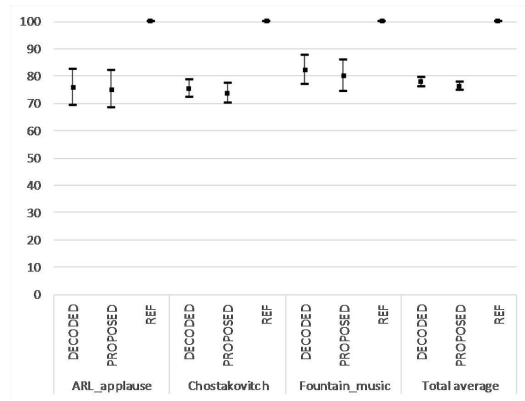


그림 8. 주관적 음질 평가 결과

Figure 8. Subjective listening test results

5. Conclusion

In this paper, we proposed the sound scene control scheme of the binaural sound in the multi-channel audio coding for generating more realistic audio sound in the VR or AR service. In the proposed method, the multi-channel gain factors and the stereo down-mix signal are

modified based on the new multi-channel layout to be re-arranged according to the user's head movement. The experimental results show that the proposed sound scene control scheme of the binaural sound in the multi-channel audio coding can successfully generate the binaural sound with the controlled sound scene while maintaining the good sound quality. Since the proposed sound scene control scheme considers the user's head movement, the azimuth control of the sound scene is only possible. As a future work, the more reliable sound scene control considering the user's free movement in the VR or AR will be studied.

References

- [1] I-J. Kim, *Trends on virtual reality technology*, Broadcastings and Media, Vol. 22, No. 2, pp. 51-60, Apr. 2016.
- [2] Y-M. Im, *Contents and trends for virtual reality technology*, Journal of the Korean Institute of Communications and Information Sciences, Vol. 33, No. 12, pp. 49-55, 2016.
- [3] K-K. Kim, and J-S. Kim, *Binaural decoding for efficient multi-channel audio service in network environment*, 2014 IEEE 11th Consumer Communications and Networking Conference, pp. 525-526, Jan. 2014.
- [4] K-K. Kim, *A study on complexity reduction of binaural decoding in multi-channel audio coding for realistic audio service*, Contemporary Engineering Sciences, Vol. 9, 2016, No. 1, pp. 11-19, Jan. 2016.
- [5] ISO/IEC 23003-1, Information Technology-MPEG Audio Technologies - Part 1: MPEG Surround, 2007.
- [6] C. Faller, and F. Baumgarte, *Binaural cue coding-Part II: Schemes and applications*, IEEE Trans. Speech Audio Processing, Vol. 11, pp. 520-531, Nov. 2003.
- [7] 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.
- [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*, Consumer Electronics, IEEE Transactions on, Vol. 51, No. 4, pp. 1253-1259, Dec. 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, Apr. 2006.
- [10] 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. on Multimedia, Vol. 13, No. 6, pp. 1208-1216, Dec. 2011.
- [11] K. Kim, J. Seo, S. Beack, D. Jang, and M. Hahn, *Improved channel level difference quantization for spatial audio coding*, ETRI Journal, Vol. 29, No. 1, pp. 99-102, Feb. 2007.
- [12] K-K. Kim, *Sound scene control of multi-channel audio signals for realistic audio service in wired/wireless network*, International Journal of Multimedia and Ubiquitous Engineering, Vol. 9, No. 2, pp. 253-258, 2014.
- [13] V. Pulkki, *Virtual sound source positioning*

using vector base amplitude panning, Journal of Audio Engineering Society, Vol. 45, pp. 456-466, Jun. 1997.

- [14] ISO/IEC JTC1/SC29/WG11 (MPEG), *Procedures for the evaluation of spatial audio coding systems*, Document N6691, Redmond, Jul. 2004.
- [15] ITU-R Recommendation, *Method for the subjective assessment of intermediate sound quality (MUSHRA)*, ITU, BS. 1543-1, Geneva, 2001.

다채널 오디오 부호화기에서의 바이노럴 사운드 썬 조절에 관한 연구

김광기

나사렛대학교 디지털콘텐츠학과

요 약

최근 3DTV, UHD TV, 3D 시네마, 3D 게임, 증강현실 (AR), 가상현실 (VR) 등과 같은 실감형 비디오 서비스가 증가하고 있으며, 실감형 비디오 서비스를 더욱 실감나게 만들 수 있는 실감형 오디오 서비스의 필요성이 커지고 있다. 일반적으로 실감형 오디오 서비스는 5.1 채널 이상의 다채널 재생 환경을 가지고 있어야 한다는 제약을 가지고 있으며, 이를 해결하기 위하여 스테레오 헤드폰 환경에서 다채널 오디오에 의한 실감형 오디오 효과를 느낄 수 있는 바이노럴 사운드의 재생 기술이 소개되었다. 그러나, 바이노럴 사운드는 사용자의 위치변화에 상관없이 항상 일정한 사운드 썬만을 가지고 있으므로 AR이나 VR과 같은 서비스에서 사용자의 위치의 변화를 반영하지 못한다는 문제를 지니고 있다. 본 논문에서는 AR이나 VR 서비스에서 더욱 실감나는 오디오 사운드 생성을 위한 바이노럴 사운드의 썬 조절 방법을 제안한다. 제안된 방법에서 다채널 오디오 신호의 이득과 스테레오 다운믹스 신호는 사용자의 위치 (방위각) 변화에 따라 재 배치되는 다채널 재생 환경을 이용하여 조정되며, 조정된 다채널 오디오 신호의 이득과 다운믹스 신호를 이용하여 사운드 썬이 조절된 바이노럴 사운드를

생성할 수 있다. 실험을 통해 제안된 방법이 음질의 열화 없이 바이노럴 사운드 썬 조절을 성공적으로 수행할 수 있음을 확인하였다.

Acknowledgments

This work was funded by the research fund of Korea Nazarene University in 2016.



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/-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