Research Article

Ultrasonic Sensor-Based Personalized Multichannel Audio Rendering for Multiview Broadcasting Services

Yong Guk Kim, Sang-Taeck Moon, Seung Ho Choi, and Hong Kook Kim

1 Maritime R&D Center, LIGNex1, 702 Sampyeong-dong, Bundang-gu, Seongnam-si, Gyeonggi 463-300, Republic of Korea
2 Department of Electronic and IT Media Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 139-743, Republic of Korea
3 School of Information and Communications, Gwangju Institute of Science and Technology (GIST), 1 Oryong-dong, Buk-gu, Gwangju 500-712, Republic of Korea

Correspondence should be addressed to Hong Kook Kim; hongkook@gist.ac.kr

Received 11 November 2012; Accepted 10 March 2013

Academic Editor: Sabah Mohammed

Copyright © 2013 Yong Guk Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An ultrasonic sensor-based personalized multichannel audio rendering method is proposed for multiview broadcasting services. Multiview broadcasting, a representative next-generation broadcasting technique, renders video image sequences captured by several stereoscopic cameras from different viewpoints. To achieve realistic multiview broadcasting, multichannel audio that is synchronized with a user’s viewpoint should be rendered in real time. For this reason, both a real-time person-tracking technique for estimating the user’s position and a multichannel audio rendering technique for virtual sound localization are necessary in order to provide realistic audio. Therefore, the proposed method is composed of two parts: a person-tracking method using ultrasonic sensors and a multichannel audio rendering method using MPEG Surround parameters. In order to evaluate the perceptual quality and localization performance of the proposed method, a MUSHRA listening test is conducted, and the directivity patterns are investigated. It is shown from these experiments that the proposed method provides better perceptual quality and localization performance than a conventional multichannel audio rendering method that also uses MPEG Surround parameters.

1. Introduction

Recently, a wide range of multimedia technologies for accessing multimedia content through digital TVs (DTVs), personal media players (PMPs), and digital cameras is rapidly being developed. This development is particularly evident in the field related to broadcasting services, which has made progress toward more realistic and immersive broadcasting services [1–5]. To this end, a representative next-generation broadcasting service that supports realistic and immersive multimedia is currently entering the spotlight in the form of 3-dimensional television (3DTV) technologies [5–7].

3DTV is a technology that is being used to provide realistic and stereoscopic video content to users and can be further classified into either stereoscopic or multiview methods. Stereoscopic 3DTV is currently being produced and sold on the market and has become an essential component for watching 3D movies at home. As an alternative to glassless 3DTV, however, multiview-based 3DTV is emerging as an attractive option, since it not only delivers more realistic visual content to users, but it also has a wider viewing range. Thus, there is a great deal of ongoing research associated with multiview TVs in attempts to miniaturize the screen size and reduce the price [7].

Multiview broadcasting renders the video sequences captured by a set of cameras from different viewpoints. By rendering these video sequences on a multiview monitor or a multiview TV, users can experience 3D effects from different viewpoints without requiring 3D glasses [7]. Under a multiview broadcasting framework, however, the transmitted multichannel audio signal must also be realistically rendered at different viewpoints in order to increase both the visual and auditory realism. To realize such an audio service, two sequential processes are necessary: (1) tracking the user’s viewpoint and (2) rendering the multichannel audio specifically at the user’s position.
Thus, this paper proposes a person-tracking-based multichannel audio rendering method for multiview broadcasting services, in which person tracking is performed using ultrasonic sensors, and multichannel audio rendering is performed using MPEG Surround parameters.

The remainder of this paper is organized as follows. Following this introduction, Section 2 briefly explains a multiview broadcasting system. Next, Section 3 proposes an ultrasonic-based person-tracking method for a personalized audio service. After that, Section 4 describes a conventional parameter-based audio rendering method and then proposes a new rendering method using MPEG Surround parameters on the basis of the constant power panning law. Section 5 then evaluates the performance of the proposed method in terms of perceptual audio quality and audio localization. Finally, this paper is concluded in Section 6.

2. Multiview Broadcasting System

Figure 1 presents a schematic diagram of a multiview and multichannel audio broadcasting system. As shown in this figure, the broadcasting system is composed of two parts: the first part acquires and transmits multiview images and multichannel audio contents, and the second part renders and plays the resultant multiview images and multichannel audio. In the first part, multiview videos consist of video sequences that are simultaneously captured by a set of cameras placed according to different viewpoints, which can then be encoded using a video encoder such as H.264. On the other hand, multichannel audio contents are recorded using multiple microphones or a microphone array, which are then encoded using an audio codec such as MPEG-2 advanced audio coding (AAC). Next, both video and audio contents are transmitted to a multiview receiver via a broadcasting network. In the second part, the transmitted multiview video contents are processed and rendered to generate 3D contents that are adjusted to the particular viewpoint of each user. Similarly, multichannel audio is rendered for each viewpoint and played through 5.1 multichannel loudspeakers or stereo headphones.

3. Ultrasonic Sensor-Based Person Tracking

In this section, we describe how the viewpoint of a user can be estimated in order to deliver audio effects appropriate to a particular viewpoint, as mentioned in Sections 1 and 2. Recently, a number of methods pertaining to person tracking have been reported [8–12], which are commonly classified into two categories: vision-based tracking and active sensor-based tracking. The former tracks a person’s eyes or face [8–12], and the latter tracks a person’s position using sensors such as an active badge, a radio frequency identification (RFID) device [11], or other sensors [12, 13]. It should be noted that vision-based tracking methods have a disadvantage in terms of processing time, since they are based on image-processing techniques. However, active sensor-based tracking methods can be implemented with less processing time than vision-based tracking methods but require sensors for estimating the viewpoint of each user. However, it has been shown that tracking methods utilizing ultrasonic devices can provide a comparatively high accuracy and are relatively inexpensive compared to RFID tags or other active badge devices [14, 15]. Consequently, in this paper, a person-tracking system using ultrasonic devices is constructed, which consists of two ultrasonic transducers and an ultrasonic receiver for person tracking.

Figure 2 presents the block diagram of a person-tracking system for estimating the user’s viewpoint, where an ultrasonic receiver attached to the user’s headphones or clothes receives an ultrasonic signal from two ultrasonic transducers. The distance between the ultrasonic receiver and each transducer is estimated and then delivered to a person-tracking server over Bluetooth. Finally, the server estimates the viewpoint using a triangulation technique.

Figure 3 shows how to calculate the view position or coordinate of the user by using the two ultrasonic sensors. The detailed procedure for person tracking is as follows. First, the relative distance between the ith ultrasonic sensor and the receiver, \( l_i \), is calculated using

\[
l_i^2 = (x_i - x_{receiver})^2 + (y_i - y_{receiver})^2 \quad \text{for } i = 1, 2 \tag{1}
\]
In this figure, panning gains in the proposed constant power panning law using MPEG Surround parameter-based audio rendering method which is based on the constant power panning law to the channel power gains as stated in Section 3, the angles to be panned are computed and denoted as \( \theta_n \) for \( n = 1, 2, \ldots, N \) (Figure 5). Note that \( N = 5 \) for a 5.1-channel speaker configuration and the angle associated with the user’s viewpoint is \( \theta_1 \). In addition, the low frequency enhancement (Lfe) channel is omitted because it can be generated by using other 5 channels. In a conventional channel power gain modification method [20, 21], the proportion of \( \phi_n \) to an aperture angle between the \( n \)th and the \((n+1)\)th channels, \( \theta_{\text{out},n} \), is calculated as

\[
\theta_{\text{out},n} = \frac{\phi_n}{\eta_{n,n+1}} \cdot \frac{\pi}{2}
\]

for \( n = 1, \ldots, N \), (5)

where CLD parameters and channel power gains are given by [16, 18]

\[
G_n = \frac{1}{\sqrt{1 + c_{n,n+1}^2}}, \quad G_{n+1} = \frac{c_{n,n+1}}{\sqrt{1 + c_{n,n+1}^2}}.
\]

(3)

where \( n \) is the channel index, and \( G_n \) and \( G_{n+1} \) are the \( n \)th and the \((n+1)\)th channel power gains, respectively. Note here that the two channels must be adjacently located. Then, if \( n \) is equal to \( N \), \( G_{n+1} \) indicates \( G_1 \), and \( c_{n,n+1} \) is a scale factor transformed from CLD using the relationship

\[
c_{n,n+1} = 10^{\text{CLD}_{n,n+1}/20},
\]

(4)

where CLD_{n,n+1} is the CLD parameter between the \( n \)th and the \((n+1)\)th channels.

Next, the channel power gains are modified depending on the panning gains calculated from a particular viewpoint, and the modified channel power gains are finally converted back into CLD parameters to create a modified bitstream for the MPEG Surround decoder.

There have been several approaches proposed for audio panning in the MPEG Surround parameter domain [19–23]. For example, the constant power panning law was directly applied to the channel power gains according to the desired panning angle [20, 21]. However, in such a direct application, the panned sound image was incorrectly localized or disappeared when the desired panning angle was larger than the aperture angle among the speaker pairs. The source of this problem was due to the fact that audio rendering coverage was limited to the aperture angles between two speakers and each transformed channel power gain was only related to two adjacent channels.

To remedy this problem, the proposed method applies the constant power panning law to the channel power gains according to the minimum aperture angle, instead of the desired panning angle. This change is especially effective when the desired panning angle is larger than any other aperture angle among the speaker pairs. In this section, a conventional channel power gain modification method in [20, 21] is reviewed, and then the proposed method is described in detail.

4.1. Conventional Channel Power Gain Modification. To track the user’s viewpoint as stated in Section 3, the angles to be panned are computed and denoted as \( \phi_n \) for \( n = 1, 2, \ldots, N \) (Figure 5). Note that \( N = 5 \) for a 5.1-channel speaker configuration and the angle associated with the user’s viewpoint is \( \phi_1 \). In addition, the low frequency enhancement (Lfe) channel is omitted because it can be generated by using other 5 channels. In a conventional channel power gain modification method [20, 21], the proportion of \( \phi_n \) to an aperture angle between the \( n \)th and the \((n+1)\)th speakers, \( \theta_{\text{out},n} \), is calculated as

\[
\theta_{\text{out},n} = \frac{\phi_n}{\eta_{n,n+1}} \cdot \frac{\pi}{2}
\]

for \( n = 1, \ldots, N \), (5)
4.2. Proposed Channel Power Gain Modification. In this section, a new audio panning method is proposed to overcome the drawbacks of the conventional method. Figure 6 shows the procedure for the proposed channel power gain modification method. In this figure, each panning angle calculated from the user’s viewpoint, $\phi_n$, is first compared to the apertures of all loudspeaker pairs, for example, five pairs of loudspeakers for the 5.1-channel speaker configuration, $\eta_{n,n+1}$ for $n = 1, 2, \ldots, 5$. Then, if the panning angle is smaller than the minimum aperture angle, the conventional method described in Section 4.1 is applied for audio panning. Otherwise, each output signal is rearranged to adjacent channels in advance before CLD panning is applied to each pair. This procedure overcomes the problem in which each channel component disappears in the output channels when the panning angle is larger than the aperture angle in sine-law amplitude panning method [24]. In other words, the output channels are arranged into another output channel

$$
\begin{align*}
\eta_{N,N+1} & = \eta_{N,1}. \\
\end{align*}
$$

Next, the panning gains associated with $\theta_{\text{out},n}$ are calculated as

$$
\begin{align*}
pG_n & = G_n \cos(\theta_{\text{out},n}), \\
pG_{n+1} & = G_{n+1} \sin(\theta_{\text{out},n}), \\
\end{align*}
$$

where $G_n$ and $pG_n$ denote the power gain of the $n$th input channel and the panning gain that is contributed from the $n$th input channel to the $(n + 1)$th speaker, respectively. In addition, the power gain of center channel is used as the panning gain of the $Lfe$ channel.

However, the conventional audio panning method described previously has some drawbacks. First, due to the sine-law amplitude panning method [24], possible panning angles in the conventional method are limited by the aperture angle of each pair of loudspeakers. Second, the conventional method does not consider the interchannel coherence (ICC) parameters for panning, though the ICC parameters play an important role in providing the spatial diffuseness of audio quality as well as localization performance at low frequencies [20].
Figure 6: Procedure of the proposed parameter-based audio panning method applied to 5.1-channel audio.

Channel power gain \( G_n \)
Panning angle \( \phi_n \)
Aperture angle of each speaker pair \( \eta_{lm,m+1} \)

\[
\phi_n > \min_{1 \leq m \leq 5} \eta_{lm,m+1} \quad \text{No}
\]
\[
\phi_{\text{remain}} = \phi_n - \min_{1 \leq m \leq 5} \eta_{lm,m+1} \quad \text{Yes}
\]

\[
\theta_{\text{out},n} = \frac{\phi_n}{\eta_{nm+1}} \cdot \frac{\pi}{2} \quad \text{for } n = 1, \ldots, 5
\]

Gain panning
\[
pG_n = G_n \cos(\theta_{\text{out},n})
\]
\[
pG_{n+1} = G_{n+1} \sin(\theta_{\text{out},n})
\]

Gain-to-CLD conversion

respective. Thus, the actual output gains of each channel are calculated as

\[
pG_{\text{out},C} = pG_{\text{out},C} + pG_{\text{L&C}}'
\]
\[
pG_{\text{out},R} = pG_{\text{out},R} + pG_{\text{R&R&S}}'
\]
\[
pG_{\text{out},Rs} = pG_{\text{out},Rs} + pG_{R\text{R&S}}'
\]
\[
pG_{\text{out},Ls} = pG_{\text{out},Ls} + pG_{L\text{S&L}}'
\]
\[
pG_{\text{out},Lfe} = pG_{\text{out},Lfe}
\]

where \( pG_{\text{out},X} \) and \( pG_{X(Y)}' \) denote the actual output gains of output channel \( X \) and the panned signal component corresponding to each speaker pair \( Y \), that is, \( (C&R), (L&C), (C&R), (R&R&S), \) and \( (R&S&L) \).

Finally, panned CLDs are obtained from both the conventional and proposed modification methods and are reestimated from the panning gains using the following equations:

\[
\text{CLD}_{0} = 20 \log_{10} \left( \frac{\sum_{X} pG_{\text{out},X}}{pG_{\text{out},C} + pG_{\text{out},Lfe}} \right)
\]
\[
\text{CLD}_{1} = 20 \log_{10} \left( \frac{pG_{\text{out},L} + pG_{\text{out},Ls}}{pG_{\text{out},R} + pG_{\text{out},Rs}} \right)
\]
\[
\text{CLD}_{2} = 20 \log_{10} \left( \frac{pG_{\text{out},C}}{pG_{\text{out},Lfe}} \right)
\]
\[
\text{CLD}_{3} = 20 \log_{10} \left( \frac{pG_{\text{out},L}}{pG_{\text{out},Ls}} \right)
\]
\[
\text{CLD}_{4} = 20 \log_{10} \left( \frac{pG_{\text{out},R}}{pG_{\text{out},Rs}} \right)
\]

where \( pG_{\text{out},X} \) denotes the channel level difference of the panned audio from the \( i \)th one-to-two (OTT) box. In addition, \( pG_{\text{out},X} \) denotes the panning gain calculated for each channel, where \( X \) is replaced with R (right channel), L (left channel), C (center channel), Rs (right surround), and Ls (left surround). Subsequently, the panned CLDs are used for MPEG Surround decoding, resulting in the panned multichannel audio shown in Figure 7 [16, 17].

5. Performance Evaluation

To evaluate the performance of the proposed audio panning method, the perceptual quality and localization performance were compared to those obtained using the conventional method. During these experiments, a multiple stimulus with hidden reference and anchor (MUSHRA) test [25] was conducted in order to evaluate the perceptual quality, and a directivity pattern analysis was used to evaluate the localization performance.
5.1. Perceptual Quality. For the MUSHRA listening test, we used the following as references and candidates: (1) a hidden reference, (2) a 7 kHz low-pass filtered anchor, (3) a 14 kHz low-pass filtered anchor, (4) audio signals processed by conventional CLD-based audio panning [20, 21], and (5) audio signals processed by the proposed CLD-based audio panning. Three music genres (classical, rock, and heavy metal) were selected as audio signals, and ten people with no hearing problems participated in these experiments.

Figure 8 illustrates the experimental results of the MUSHRA test. When the panning angle was smaller than the minimum aperture angle, for example, at a 30° panning angle, the proposed method had audio quality comparable to the conventional method, except for classical music signals. The reason why the MUSHRA score for classical music signals processed by the conventional CLD-based panning method was slightly higher than that by the proposed CLD-based panning method was that classical music signals were less dynamic than those from other genres such as rock and heavy metal. In other words, while the conventional method computed panning gains once every pair of channels by applying (6), the proposed method computed each panning gain by taking into account more than two channels as shown in (10). Thus, it resulted in perceptual degradation in classical music signals. In spite of such an artifact, it was found that the spatial impression for panned audio processed by the proposed method was more stable than that by the conventional method.

On the other hand, when the panning angle was larger than the minimum aperture angle, for example, at a 60° panning angle with a 30° minimum aperture, the audio quality of the panned audio processed by the conventional method notably degraded. Even if the proposed method had smaller MUSHRA score for classical music signals than the conventional method, it was also found that the participants heard unnatural artificial noise due to incorrect panning when the panning angle was larger than the minimum aperture.

Figure 7: Structure of MPEG Surround decoding tree.

Figure 8: MUSHRA test results at panning angles of (a) 30° and (b) 60°.
5.2. Localization Performance. To evaluate the localization performance, panned audio with only one channel signal was played, and the frequency response was measured using a dummy head. The directivity patterns for panning angles of 0°, 30°, and 60° were then analyzed. The amplitudes of the frequency responses at 500 Hz were measured by rotating the dummy head about 10°. For this experiment, a KU100 dummy head [26] was used.

Figure 9 shows the directivity patterns of the panned signals for 30° and 60° at 500 Hz. To estimate the position of the sound image localization, it was assumed that the sound image was localized at the position exhibiting maximum...
power. As illustrated in this figure, the measured power became maximal at a rotated position of about 90°, which corresponds to a forward-facing direction when no audio panning was applied. Similarly, the measured power became maximal at a rotated position of about 120°, relative to the panned direction, when an audio panning of 30° was applied. It can also be seen that the directivity pattern of the conventional method is correctly presented for a panning angle of 30°. However, when the panned angle was increased to 60°, the polar pattern of the conventional method was not correctly presented, whereas the directivity pattern obtained by the proposed CLD-based panning method shows that the audio signal rotated in the correct direction, although there were localization errors at around 5°–10°.

6. Conclusion

In this paper, an ultrasonic sensor-based personalized multichannel audio rendering method was proposed to increase audio realism in multiview broadcasting services. To this end, a real-time person-tracking method was first developed by using two ultrasonic transducers and an ultrasonic receiver in order to estimate the viewpoint of a user. Secondly, a parameter-based audio panning method using MPEG Surround parameters was proposed to increase the auditory realism. In the proposed method, panning gains were calculated according to the user’s viewpoint that was already estimated by the ultrasonic-based person-tracking method. Next, five different channel level difference (CLD) parameters were extracted from the audio bitstream after applying a CLD parser. Finally, the CLD parameters were transformed into six channel power gains for the 5.1 audio channels. In fact, the proposed method applied the constant power panning law to the channel power gains according to the minimum aperture angle, instead of the desired panning angle that was used for a conventional panning method. Thus, the proposed method could be more effective than the conventional method when the desired panning angle was larger than any other aperture angle among the speaker pairs. In order to evaluate the performance of the proposed audio panning method, the perceptual quality and localization performance using an MUSHRA test and a directivity pattern analysis, respectively, were carried out. Consequently, it was shown from the tests that the proposed audio panning method achieved better average MUSHRA score and localization performance than the conventional audio panning method.

Acknowledgment

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MEST) (no. 2012-010636).

References

[1] Y. Shishikui, Y. Fujita, and K. Kubota, “Super HD-vision demos at IBC-2008—NHK,” EBU Technical Review, January 2009.
[2] S. Y. Kim, S. U. Yoon, and Y. S. Ho, “Realistic broadcasting using multi-modal immersive media,” in Advances in Multimedia Information Processing—PCM 2005, vol. 3768 of Lecture Notes in Computer Science, pp. 164–175, 2005.
[3] K. Mitani, M. Kanazawa, K. Hamasaki, Y. Nishida, K. Shogen, and M. Sugawara, “Current status of studies on ultra high definition television,” SMPTE Motion Imaging Journal, vol. 116, no. 9, pp. 377–381, 2007.
[4] K. Hamasaki, K. Hiyama, and R. Okumura, “The 22.2 multichannel sound system and its application,” in Proceedings of the 118th AES Convention, Barcelona, Spain, May 2005, preprint 6406.
[5] A. Ando, K. Hamasaki, A. Imai et al., “Production and live transmission of 22.2 multichannel sound with ultra-high definition TV,” in Proceedings of the 122nd AES Convention, Vienna, Austria, May 2007, preprint 7137.
[6] C. Fehn, P. Kauff, M. op de Beeck et al., “An evolutionary and optimised approach on 3D-TV,” in Proceedings of International Broadcast Conference, pp. 357–365, Amsterdam, The Netherlands, September 2002.
[7] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntiens, “A survey of perceptual evaluations and requirements of three-dimensional TV,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 3, pp. 381–391, 2004.
[8] M. le Cascia, S. Sclaroff, and V. Athitsos, “Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3D models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 4, pp. 322–336, 2000.
[9] P. Viola and M. J. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
[10] M. Andersen, R. S. Andersen, N. Katsarakis, A. Pnevmatikakis, and Z. H. Tan, “Three-dimensional adaptive sensing of people in a multi-camera setup,” in Proceedings of the European Signal Processing Conference (EUSIPCO ’10), pp. 964–968, Aalborg, Denmark, August 2010.
[11] R. Want and A. Hopper, “Active badges and personal interactive computing objects,” IEEE Transactions on Consumer Electronics, vol. 38, no. 1, pp. 10–20, 1992.
[12] V. K. Singh, H. Lim, R. Malliyae, and W. Y. Chung, “Passive and cost effective people location tracking system for indoor environments using distributed wireless sensor network,” in Proceedings of World Congress on Medical Physics and Biomedical Engineering, vol. 14, pp. 392–395, Seoul, Korea, September 2006.
[13] C. H. Ou, K. F. Ssu, and H. C. Jiau, “Range-free localization with aerial anchors in wireless sensor networks,” International Journal of Distributed Sensor Networks, vol. 2, no. 1, pp. 1–21, 2006.
[14] H. Koyuncu and S. H. Yang, “A survey of indoor positioning and object locating systems,” International Journal of Computer Science and Network Security, vol. 10, no. 5, pp. 121–128, 2010.
[15] L. M. Li, Y. Liu, Y. C. Lau, and A. P. Patil, “LANDMARC: indoor location sensing using active RFID,” Wireless Networks, vol. 10, no. 6, pp. 701–710, 2004.
[16] ISO/IEC FDIS 23003-1:2006(E), MPEG Audio Technologies—Part1: MPEG Surround, 2004.
[17] J. Breebaart, L. Villemoes, and K. Kjørbing, “Binaural rendering in MPEG surround,” EURASIP Journal on Advances in Signal Processing, vol. 2008, Article ID 732895, 14 pages, 2008.
[18] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, “Parametric coding of stereo audio,” EURASIP Journal on Applied Signal Processing, vol. 2005, no. 9, pp. 1305–1322, 2005.
[19] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, “Parametric coding for high-quality audio,” in Proceedings of the 114th AES Convention, Amsterdam, Netherlands, March 2003, Preprint 5852.

[20] S. Baeck, J. Seo, I. Jang, and D. Y. Jang, “Multichannel sound scene control for MPEG surround,” in Proceedings of the 29th AES International Conference: Audio for Mobile and Handheld Devices, pp. 63–66, Seoul, Korea, September 2006.

[21] S. Baeck, J. Seo, T. Lee, and D. Y. Jang, “Spatial cue based sound scene control for MPEG surround,” in Proceedings of IEEE International Conference on Multimedia and Expo (ICME '07), pp. 1886–1889, Beijing, China, July 2007.

[22] B. Cheng, C. Ritz, and I. Burnett, “Squeezing the auditory space: a new approach to multi-channel audio coding,” in Proceedings of the 7th Pacific Rim Conference on Advances in Multimedia Information Processing (PCM '06), vol. 4261 of Lecture Notes in Computer Science, pp. 572–581, November 2006.

[23] S. J. Choi, Y. W. Jung, H. J. Kim, and H. O. Oh, “New CLD quantization method for spatial audio coding,” in Proceedings of the 120th AES Convention, Paris, France, May 2006, Preprint 6734.

[24] B. B. Bauer, “Phasor analysis of some stereophonic phenomena,” Journal of Acoustic Society of America, vol. 33, no. 11, pp. 1536–1539, 1961.

[25] ITU-R Recommendation BS. 1534-1, Method for the Subjective Assessment of Intermediate Quality Levels of Coding System, January 2003.

[26] Georg Neumann GmbH, Product Information KU 100, Berlin, Germany, November 2000.