2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), Bandung, Indonesia, October 29–30, 2015

R-TTT Module with Modified Residual Signal for Improving Multichannel Audio Signal Accuracy

Ikhwana Elfitri, Amirul Luthfi, and Fitrilina
Department of Electrical Engineering, Faculty of Engineering, Andalas University,
Kampus UNAND Limau Manih, Padang, 25163, Sumatera Barat, Indonesia
E-mail: ikhwana@ft.unand.ac.id

***Abstract*—Spatial audio coding is a technique that capable of representing multichannel audio signals as a lower number of audio channels accompanied by spatial parameters and residual signal which will be useful for recreating the original multichannel audio signals. Moving Picture Expert Group (MPEG) Surround, an international standard developed based on spatial audio coding, specifies Reverse Two-To-Three (R-TTT) module to extend stereo audio, consisted of left and right channels, into three audio channels: left, centre, and right channels based on Channel Prediction Coefficient (CPC) as spatial parameter and residual signal. In this paper, a modified residual signal is proposed to provide a better audio waveform reconstruction in the decoder side by minimising distortion caused by quantisation of CPC. Our experiments show that the waveform accuracy in terms of Signal-to-Noise Ratio (SNR) gets improved as high as 11 dB while the subjective test shows that the proposed method does not reduce perceptual quality, in terms of Subjective Difference Grade (SDG) score, of the reconstructed audio signals.**

***Keywords*—MPEG Surround, Spatial Audio Coding, Multichannel Audio Signals**

I. Introduction

Ultra High Definition Television (UHDTV), widely known as Super Hi-Vision [1]–[3], has been announced as the International Telecommunication Union (ITU) standard for the future TV broadcasting where the quality of rendered audio and video are several times higher than current advance systems. It is expected that users when using UHDTV will be able to sense much more realistic audiovisual as well as to increase the feeling of presence. While the video resolution will be much increased, the audio part of the UHDTV will employ a 22.2 loudspeaker configuration [4] to provide users with a three dimensional (3D) audio or spatial audio [5] experience.

In order to deliver 3D audio rendering, multichannel audio signal processing, ranging from recording, coding, and reproduction, is fundamental. The conventional multichannel audio configuration is 5.1 system while the UHDTV standard offers much more advance system than this traditional configuration. Moreover, the UHDTV standard offers possibility to apply object-based audio system where users can interact in the rendering system to change and adjust the audio composition. For transmission and compression of this 3D audio, a compression standard has also been introduced by Moving Picture Expert Group (MPEG), named MPEG-H 3D Audio [6]. Even though other processing chains are also very important, however, multichannel audio coding can be considered as the most essential part due to the possibility of transmitting much lower audio data particularly for channel number as many as 22 as in the case of UHDTV.

With regard to compression approach, spatial audio coding [7]–[13] has been very popular because of its capability to represent multichannel audio signal as low number of channel as possible, normally a single (mono) or dual (stereo) channels. These reduced channels must be accompanied by spatial parameters and additional residual signal so that it will be possible to recreate back multichannel configuration based on the received mono or stereo along with the decoded spatial parameters and the residual signal. Parametric Stereo [14]–[16] and MPEG Surround [17]–[20] are among the earliest MPEG standards that are developed with the basis on the spatial audio coding concept. MPEG Surround allows transmission of multichannel audio signal with operating bitrates as low as 64 and 96 kb/s for 5.1 audio system. These bitrates are much lower than what can be achieved by traditional systems such as Advanced Audio Coding (AAC) [21], [22] which typically operates at 320 kb/s.

More importantly, MPEG Surround can be considered as having placed a framework for multichannel audio compression. This can be justified in three reasons. First, MPEG Surround offers bitrate scalability making it possible to deliver multichannel audio content with different bitrates depending on the network environment. Second, it is conceptually a backward compatible system which can be implemented gradually to upgrade any existing mono or stereo system to multichannel system. Third, it is much more interesting when considering apply other functionalities, such as binaural rendering [23] for rendering multichannel audio content by a headset and artistic downmix signal for maintaing the best possible audio mono or stereo for existing users. MPEG has developed the subsequent standard such as MPEG Spatial Audio Object Coding [24], [25] and MPEG-H 3D Audio based on this framework.

One efficient module specified in MPEG Surround is Reverse Two-To-Three (R-TTT) module which reduced three audio channels: the left, centre and right channels into stereo audio channels. To make the channels reconstructable Channel Prediction Coefficient (CPC), which is basically motivated by multichannel prediction technique [26], is extracted. Furthermore, residual signal is also transmitted in order to make the prediction more accurate. In this paper, a modified R-TTT module is proposed which is capable of increasing the accuracy of reconstructed audio signals by minimising the distortion introduced during the quantisation process. This modified method to generate residual signal can also be considered as a closed-loop approach where error minimisation is applied as opposed to open-loop approach specified in the standard R-TTT module.

978-1-4673-7408-8/15/$31.00

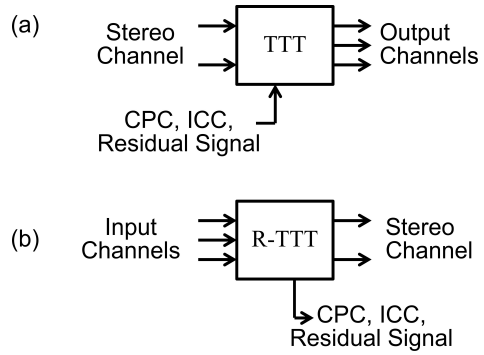

Fig. 1. Block diagram of (a) the Two-To-Three (TTT) and (b) the Reverse-TTT (R-TTT) modules. The TTT is a module specified in MPEG Surround standard to recreate three channels: left (L), centre (C), and right (R) from stereo channels while R-TTT module can be used to perform the reverse process.

Following this introduction, the other section in this paper is organised as follows. An overview of R-TTT module, as a basis for the proposed method, is presented in Section II. Then, Section III will discussed in details the proposed modified residual signal after presentation of distortion analysis introduced in the signals and parameters transmitted from the R-TTT module. Our experiments and the results to show the performance of the proposed system are discussed in Section IV while the conclusion of this work is given in Section V.

II. Overview of The R-TTT Module

MPEG Surround specifies two basic modules to reconstruct multichannel audio signals from available mono or stereo audio signals based on the decoded spatial parameters. The first module called One-To-Two (OTT) is used to reproduce two channels from a single audio channel while its corresponding module named Reverse OTT (R-OTT) is applied to do the reverse process creating a single audio channel from two available channels. Together with TTT and R-TTT modules, both OTT and R-OTT modules are employed to downmix multichannel audio signals into a mono or stereo audio or in oppose to correspondingly recreate multichannel audio signals from a mono or stereo audio. Both pairs of modules i.e. R-OTT and OTT as well as R-TTT and TTT, are performed in subband domain of hybrid filterbank which decompose audio signals into 71 hybrid bands in order to process the audio signals in a way that is similar to the process in the human hearing system. In this section, it is only the TTT and R-TTT modules which will be discussed in more details. However, for more details on MPEG Surround as well as the OTT and R-OTT modules, it is recommended to refer to [27].

The TTT and R-TTT modules are shown in Fig. 1. The TTT module is designed to convert a stereo audio (two channels) to three audio channels. Reciprocally, the R-TTT module is used to convert three audio channels into a stereo audio channel. Two encoding modes are available: the prediction and the energy modes. Using the prediction mode, the Channel Prediction Coefficient (CPC) along with residual signal are calculated as spatial parameter and transmitted as side information while using the energy mode the residual signal can be replaced by the Inter-Channel Coherence (ICC).

Let the left, right, and center channels, $x_L[n], x_R[n], x_C[n]$, be the input channels of the R-TTT module while the other left and the right channel, $y_L[n], y_R[n]$, are the output stereo channels. The two outputs, along with an auxiliary signal, $y_C[n]$, can be represented as linear combinations of the input signals as below

$$y_L[n] = x_L[n] + \tfrac{1}{2}\sqrt{2}\ x_C[n], \tag{1a}$$
$$y_R[n] = x_R[n] + \tfrac{1}{2}\sqrt{2}\ x_C[n], \tag{1b}$$
$$y_C[n] = x_L[n] + x_R[n] - \tfrac{1}{2}\sqrt{2}\ x_C[n]. \tag{1c}$$

When the prediction mode is used, two CPC coefficients are calculated in such a way that the auxiliary signal, $y_C[n]$, is as close as the linear combination of the two outputs according to

$$\hat{y}_C[n] = \gamma_1\ y_L[n] + \gamma_2\ y_R[n]. \tag{2}$$

Based on these linear combinations the CPC, γ, can be derived as follow:

$$\gamma_1 = \frac{< \mathbf{y}_L, \mathbf{y}_C >^* ||\mathbf{y}_R||^2 - < \mathbf{y}_R, \mathbf{y}_C >^* < \mathbf{y}_L, \mathbf{y}_R >^*}{||\mathbf{y}_L||^2||\mathbf{y}_R||^2 - | < \mathbf{y}_L, \mathbf{y}_R > |^2} \tag{3}$$

$$\gamma_2 = \frac{< \mathbf{y}_R, \mathbf{y}_C >^* ||\mathbf{y}_L||^2 - < \mathbf{y}_L, \mathbf{y}_C >^* < \mathbf{y}_L, \mathbf{y}_R >^*}{||\mathbf{y}_L||^2||\mathbf{y}_R||^2 - | < \mathbf{y}_L, \mathbf{y}_R > |^2} \tag{4}$$

where

$$< \mathbf{y}_{ch1}, \mathbf{y}_{ch2} > \equiv \sum_n y_{ch1}[n] \cdot y^*_{ch2}[n] \tag{5}$$

$$||\mathbf{y}_{ch1}||^2 \equiv \sum_n |y_{ch1}[n]|^2 \tag{6}$$

The residual signal is defined as the difference between the auxiliary signal, $y_C[n]$, and the predicted auxiliary signal, $\hat{y}_C[n]$, as

$$r_C[n] = y_C[n] - \hat{y}_C[n]. \tag{7}$$

In case the residual signal is not transmitted to the decoder (this is the case for low bitrate transmission) the corresponding energy loss can be described by transmitting the Inter-Channel Coherence (ICC).

Using the prediction mode, a reliable estimation of the auxiliary signal is required for the decoder. If it is difficult to guarantee the reliability of the prediction mode, then the TTT using an energy mode can be applied. The energy mode is performed by describing the relative energy distribution among the three input channels. Two CLD parameters can be transmitted. The prediction and energy mode can be performed independently for each parameter band.

In the decoding process using the prediction mode, the TTT module is employed to recreate three channels from a stereo channel as below

$$\hat{x}_L[n] = \tfrac{2}{3}\hat{y}_L[n] - \tfrac{1}{3}\hat{y}_R[n] + \tfrac{1}{3}y_C[n], \tag{8a}$$
$$\hat{x}_R[n] = -\tfrac{1}{3}\hat{y}_L[n] + \tfrac{2}{3}\hat{y}_R[n] + \tfrac{1}{3}y_C[n], \tag{8b}$$
$$\hat{x}_C[n] = \tfrac{1}{3}\sqrt{2}\hat{y}_L[n] + \tfrac{1}{3}\sqrt{2}\hat{y}_R[n] - \tfrac{1}{3}\sqrt{2}y_C[n]. \tag{8c}$$

where the predicted auxiliary signal is calculated from the CPCs and the residual signal as in (2) and (7). If the residual signal is not available (i.e. not transmitted from the encoder side), the TTT module has two options to recreate the three

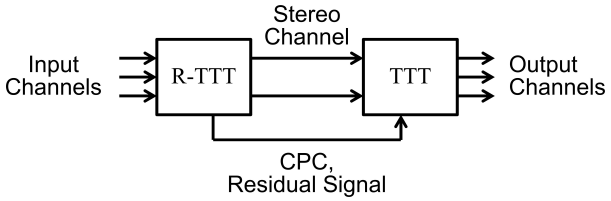

Fig. 2. Illustration of the proposed Modified R-TTT module which can also be considered as a closed-loop configuration for minimising distortion.

outputs. The first one is to apply a suitable gain to each output based on the CPC and ICC as below

$$\hat{x}_L[n] = \left(\frac{\gamma_1 + 2}{3\text{ICC}}\right) y_L[n] - \left(\frac{\gamma_2 - 1}{3\text{ICC}}\right) y_R[n], \tag{9a}$$

$$\hat{x}_R[n] = \left(\frac{\gamma_1 - 1}{3\text{ICC}}\right) y_L[n] - \left(\frac{\gamma_2 + 2}{3\text{ICC}}\right) y_R[n], \tag{9b}$$

$$\hat{x}_C[n] = \left(\frac{\sqrt{2}(1-\gamma_1)}{3\text{ICC}}\right) y_L[n] - \left(\frac{\sqrt{2}(1-\gamma_2)}{3\text{ICC}}\right) y_R[n]. \tag{9c}$$

The second is to create a synthetic residual signal by means of a decorrelator.

The extracted CPC parameters need to be digitised with a uniform quantiser specified in MPEG Surround standard. The stepsize of the quantiser is 0.1 with the maximum quantisation value is $+3.0$ while the minimum quantisation value is -2.0.

III. The Proposed R-TTT Module with Modified Residual Signal

The standard R-TTT module is considered as having no method to minimise distortion. In this section the distortion caused by the R-TTT module is discussed first, followed by the details on the proposed modified R-TTT module. It is important to note that the proposed scheme is only aimed to work in the prediction mode.

A. Analysis on Distortion of R-TTT Module

Let us start by trying to find the representation of the recreated left, right, and centre channel audio signals in the TTT module from (8) by substituting (1). In a hypothetical condition where there is no error at all signals and parameters received by the TTT module (please refer to an illustration in Fig. 2), the received left channel signal as output of the TTT module will be identical to the original left channel signal which can be shown as below:

$$\begin{aligned}\hat{x}_L[n] = {} & \tfrac{2}{3}x_L[n] + \tfrac{1}{3}x_L[n] - \tfrac{1}{3}x_R[n] + \tfrac{1}{3}x_R[n] \\ & -\tfrac{1}{6}\sqrt{2}\ x_C[n] + \tfrac{1}{3}\sqrt{2}\ x_C[n] - \tfrac{1}{6}\sqrt{2}\ x_C[n]\end{aligned} \tag{10}$$

hence, $\hat{x}_L[n] = x_L[n]$. Using the same way, we will have the reproduced right and centre channel signals be equal to the original ones, $\hat{x}_R[n] = x_R[n]$ and $\hat{x}_C[n] = x_C[n]$. However, in the real transmission scenario, there are at least three kinds of distortion due to coding and quantisation processes: coding errors in the received stereo and residual signals as well as quantisation error in the received CPC parameters. Another prediction error actually also exists in a situation where the predicted auxiliary signal differs to the original one. However, the R-TTT module has been equipped with residual coding method to compensate for this prediction error.

Among those distortion there is a type of error that can be minimised which is the error due to the quantisation of the CPC parameter. We can elaborate further this distortion by extending the previous assumption. In the case of the CPC parameter is quantised and transmitted to the decoder, the predicted auxiliary signal at the decoder side can be written as below:

$$y_C[n] = \hat{r}_C[n] + \hat{\gamma}_1\ \hat{y}_L[n] + \hat{\gamma}_2\ \hat{y}_R[n] \tag{11}$$

and then the left channel can be reconstructed by substituting (11) to (8), resulting in as follow:

$$\begin{aligned}\hat{x}_L[n] = {} & \frac{\hat{\gamma}_1 + 2}{3}x_L[n] + \frac{\hat{\gamma}_2 - 1}{3}x_R[n] \\ & + \frac{1}{3}\hat{r}_C + \frac{1 + \hat{\gamma}_1 + \hat{\gamma}_2}{6}\sqrt{2}\ x_C[n]\end{aligned} \tag{12}$$

Following the same way, the other channels, the right channel signal can be recreated as below:

$$\begin{aligned}\hat{x}_R[n] = {} & \frac{\hat{\gamma}_1 - 1}{3}x_L[n] + \frac{\hat{\gamma}_2 + 2}{3}x_R[n] \\ & + \frac{1}{3}\hat{r}_C + \frac{1 + \hat{\gamma}_1 + \hat{\gamma}_2}{6}\sqrt{2}\ x_C[n]\end{aligned} \tag{13}$$

while the centre channel signal can be reproduced as:

$$\begin{aligned}\hat{x}_C[n] = {} & \frac{\sqrt{2}(1-\hat{\gamma}_1)}{3}x_L[n] + \frac{\sqrt{2}(1-\hat{\gamma}_2)}{3}x_R[n] \\ & - \frac{\sqrt{2}}{3}\hat{r}_C + \frac{2 - \hat{\gamma}_1 - \hat{\gamma}_2}{3}\ x_C[n]\end{aligned} \tag{14}$$

B. Modified Residual Signal

In order to provide a mechanism of minimising distortion due to quantisation of CPC parameters, we employ a modified method to determine residual signal as follow:

$$\begin{aligned}r_C[n] = {} & (1 - \hat{\gamma}_1)\ x_L[n] + (1 - \hat{\gamma}_2)\ x_R[n] \\ & -\tfrac{1}{2}\sqrt{2}(1 + \hat{\gamma}_1 + \hat{\gamma}_2)\ x_C[n]\end{aligned} \tag{15}$$

which is derived based on the predicted auxiliary signal as below:

$$\hat{y}_C[n] = \hat{\gamma}_1\ y_L[n] + \hat{\gamma}_2\ y_R[n]. \tag{16}$$

Transmitting the modified residual signal to the decoder is capable of reproducing three channels audio signal with minimised distortion. For instance in the case of only quantisation error exists, we can show that the reconstructed left channel audio signal is equal to the original left channel signal. This can be shown by substitution of (16) and (1) to (12), resulting in the following representation:

$$\begin{aligned}\hat{x}_L[n] = {} & \frac{\hat{\gamma}_1 + 2}{3}x_L[n] + \frac{1 - \hat{\gamma}_1}{3}x_L[n] \\ & + \frac{\hat{\gamma}_2 - 1}{3}x_R[n] + \frac{1 - \hat{\gamma}_2}{3}x_R[n] \\ & + \frac{1 + \hat{\gamma}_1 + \hat{\gamma}_2}{6}\sqrt{2}\ x_C[n] \\ & - \frac{1 + \hat{\gamma}_1 + \hat{\gamma}_2}{6}\sqrt{2}\ x_C[n]\end{aligned} \tag{17}$$

TABLE I. SNR COMPARISON (IN DECIBELS) BETWEEN THE STANDARD AND PROPOSED R-TTT

Audio Excerpt	Standard R-TTT	Proposed R-TTT
clap	30.86	49.74
drum	24.92	34.86
lough	27.71	39.02
news	27.04	37.74
music	25.29	29.73
Mean	27.16	38.22

where we can easily see that $\hat{x}_L[n] = x_L[n]$. We can also apply the corresponding substitution to generate representation for the other channel as follows:

$$\begin{aligned}\hat{x}_R[n] = \frac{\hat{\gamma}_1 - 1}{3} x_L[n] + \frac{1-\hat{\gamma}_1}{3} x_L[n] \\ + \frac{\hat{\gamma}_2 + 2}{3} x_R[n] + \frac{1-\hat{\gamma}_2}{3} x_R[n] \\ + \frac{1+\hat{\gamma}_1+\hat{\gamma}_2}{6}\sqrt{2}\; x_C[n] \\ - \frac{1+\hat{\gamma}_1+\hat{\gamma}_2}{6}\sqrt{2}\; x_C[n]\end{aligned} \tag{18}$$

as well as,

$$\begin{aligned}\hat{x}_C[n] = \frac{\sqrt{2}(1-\hat{\gamma}_1)}{3} x_L[n] - \frac{\sqrt{2}(1-\hat{\gamma}_1)}{3} x_L[n] \\ + \frac{\sqrt{2}(1-\hat{\gamma}_2)}{3} x_R[n] - \frac{\sqrt{2}(1-\hat{\gamma}_2)}{3} x_R[n] \\ + \frac{2-\hat{\gamma}_1-\hat{\gamma}_2}{3}\; x_C[n] \\ + \frac{1+\hat{\gamma}_1+\hat{\gamma}_2}{3}\sqrt{2}\; x_C[n]\end{aligned} \tag{19}$$

which yield $\hat{x}_R[n] = x_R[n]$ and $\hat{x}_C[n] = x_C[n]$.

For a better comparison we can derive expressions for the reconstructed audio signals on the TTT module when conventional residual signal as in (2) is applied and substituted in (12). Hence, the reconstructed signal on the left channel can be written as below:

$$\begin{aligned}\hat{x}_L[n] = x_L[n] + \frac{\hat{\gamma}_1 - \gamma_1}{3} x_L[n] + \frac{\hat{\gamma}_2 - \gamma_2}{3} x_R[n] \\ + \frac{\hat{\gamma}_1 - \gamma_1 + \hat{\gamma}_2 - \gamma_2}{6}\sqrt{2}\; x_C[n]\end{aligned} \tag{20}$$

The representation for the right channel can be determined based on the same way which results in as follow:

$$\begin{aligned}\hat{x}_R[n] = \frac{\hat{\gamma}_2 - \gamma_2}{3} x_L[n] + x_R[n] + \frac{\hat{\gamma}_1 - \gamma_1}{3} x_R[n] \\ + \frac{\hat{\gamma}_1 - \gamma_1 + \hat{\gamma}_2 - \gamma_2}{6}\sqrt{2}\; x_C[n]\end{aligned} \tag{21}$$

while the reconstructed signal on the centre channel can be written as below:

$$\begin{aligned}\hat{x}_C[n] = \frac{\hat{\gamma}_2 - \gamma_2}{3} x_L[n] + \frac{\hat{\gamma}_1 - \gamma_1}{3} x_R[n] \\ + x_C[n] + \frac{\hat{\gamma}_1 - \gamma_1 + \hat{\gamma}_2 - \gamma_2}{3}\sqrt{2}\; x_C[n]\end{aligned} \tag{22}$$

The comparison of applying the conventional and modified residual signals shows that when conventional residual signal is applied, quantisation error reflected on the difference between CPC, γ, and quantised CPC, $\hat{\gamma}$, can be easily noticed on every reproduced signal. However, when modified residual signal is employed the quantisation error can be minimised.

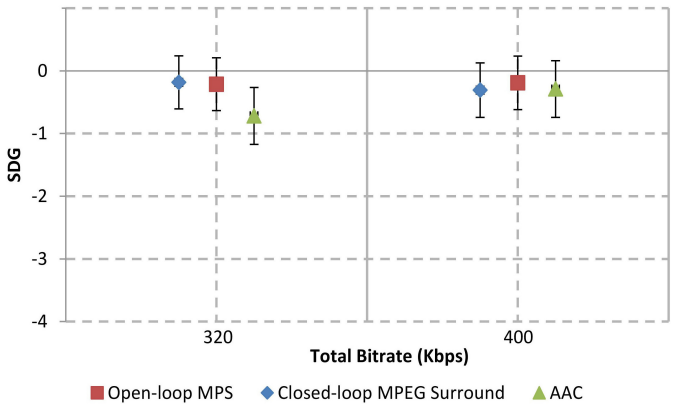

Fig. 3. Subjective DIfference Grade (SDG) scores of the proposed Modified R-TTT module (closed-loop MPEG Surround) in comparison with the standard R-TTT module (open-loop MPS) and Advanced Audio Coding (AAC) multichannel.

IV. EXPERIMENTAL RESULTS

The proposed R-TTT module has been evaluated particularly in its performance to improve signal accuracy in terms of Signal-to-Noise Ratio (SNR) and then subjective test was done to verify that the proposed module does not reduce the perceptual quality. Five audio excerpts as listed in Table I, each of them is 5-channel audio signals, were used as the audio input for the experiments. Table I presents SNR measurements of the standard and the proposed R-TTT modules. For these experiments the continuous down-mix and residual signals without further encoding were used while only the spatial parameters were quantised. The results show that for all audio excerpts the SNR achieved by the proposed module are considerably higher than the standard R-TTT module even though the improvement is different for each audio excerpt. In average, the proposed module can increase the SNR approximately 11 dB which indicates that signal accuracy is improved.

The MPS encoder, using modified residual signal and operating at 2 total bitrates: 320 and 400 kb/s, were also subjectively assessed based on ITU-R BS.111-6 Recommendation [28]. The 5-channel input signals were down-mixed into stereo channels which are then encoded by Advanced Audio Coding (AAC). For comparison, the MPS using standard R-TTT module and AAC multichannel were included in the listening test. The results are given in Fig. 3 as average Subjective Difference Grade (SDG) score ranging from 0 to -4 presented with 95 percent confidence interval. The best quality is graded as 0 which means the audio artifact is imperceptible. Grade -1 is used for audio quality with perceptible artifact but not annoying. Grade -2 to -4 indicate the artifacts are: slightly annoying, annoying, and very annoying, respectively. As many as 21 experienced subjects participated in the listening test. During postscreening greater-than-zero SDG scores, which means invalid, from 6 subjects were discarded. The results demonstrate that the average SDG from all three codecs: standard R-TTT module (open-loop MPS), modified R-TTT module (Closed-loop MPEG Surroud), and AAC, are

statistically comparable. It indicates that all three tested audio codecs are considered as competitive and the proposed R-TTT module does not reduce perceptual quality of the tested audio signals.

V. Conclusion

A modified residual signal has been proposed in this paper which is capable of minimising distortion caused by the quantisation of the CPC parameters. The proposed technique has been shown to be able to increase the audio waveform accuracy in terms of SNR while maintaining the perceptual quality, in terms of SDG score, of the reconstructed multichannel audio signals. For a better understanding on the proposed modified residual signal technique, analysis on the distortion introduced on the R-TTT module has also been presented.

Acknowledgment

This work was funded by the Ministry of Research, Technology and Higher Education, the Republic of Indonesia under the scheme of Higher Education Excellent Research (Penelitian Unggulan Perguruan Tinggi) with contract no. 50/UN.16/UPT/LPPM/2015. The authors would like to thank Dr. Rahmadi Kurnia and Mr. Heru Dibyo Laksono for meaningful suggestions during conducting the experiments and preparing this manuscript. Furthermore, the authors also thank the reviewers for their constructive comments and suggestions.

References

[1] E. Nakasu, "Super Hi-Vision on the horizon: A future TV system that conveys an enhanced sense of reality and presence ," *IEEE Consumer Electronics Magazine*, vol. 1, no. 2, pp. 36–42, March 2012.

[2] Y. Shishikui, K. Iguchi, S. Sakaida, K. Kazui, and A. Nakagawa, "High-performance video-codec for Super Hi-Vision," *Proceedings of the IEEE*, vol. 101, no. 1, pp. 130–139, January 2013.

[3] T. Ito, "Future television - super hi-vision and beyond," in *Proc. IEEE Asian Solid State Circuits Conference*, Beijing, China, Nov. 2010.

[4] T. Sugimoto, Y. Nakayama, and S. Oode, "Bitrate of 22.2 multichannel sound signal meeting broadcast quality," in *Proc. 137th AES Convention*, Los Angeles, USA, Oct. 2014.

[5] K. Brandenburg, C. Faller, J. Herre, J. D. Johnston, and W. B. Kleijn, "Perceptual coding of high-quality digital audio," *Proceedings of the IEEE*, vol. 101 No. 9, pp. 1905–1919, 2014.

[6] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio The new standard for universal spatial/3d audio coding," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 821–830, 2015.

[7] J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Hoelzer, K. Linzmeier, C. Spenger, and P. Kroon, "Spatial audio coding: Next-generation efficient and compatible coding of multi-channel audio," in *Proc. the* 117^{th} *Convention of the Audio Engineering Society*, San Fransisco, CA, USA, Oct. 2004.

[8] J. Herre, "From joint stereo to spatial audio coding - recent progress and standardization," in *Proc. of the* 7^{th} *Int. Conf. on Digital Audio Effects (DAFx'04)*, Naples, Italy, October 2004.

[9] J. Herre and S. Disch, "New concepts in parametric coding of spatial audio: From SAC to SAOC," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, San Fransisco, CA, USA, Oct. 2007.

[10] I. Elfitri, B. Gunel, and A. M. Kondoz, "Multichannel audio coding based on analysis by synthesis," *Proc. of the IEEE*, vol. 99, no. 4, pp. 657–670, April 2011.

[11] I. Elfitri, X. Shi, and A. M. Kondoz, "Analysis by synthesis spatial audio coding," *IET Signal Processing*, vol. 8, no. 1, pp. 30–38, February 2014.

[12] I. Elfitri, R. Kurnia, and Fitrilina, "Investigation on objective performance of closed-loop spatial audio coding," in *Proc. of 2014 Int. Conf. on Information Tech. and Electrical Eng.*, Jogjakarta, Indonesia, October 2014.

[13] I. Elfitri, A. Permana, and H. D. Laksono, "Balanced-delay filterbank for closed-loop spatial audio coding," in *Proc. of 2015 Int. Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia, May 2015.

[14] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 1305–1322, 2005.

[15] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, "Low complexity parametric stereo coding," Presented at the 116^{th} Convention of the Audio Engineering Society, Berlin, Germany, May 2004.

[16] I. Elfitri, R. Kurnia, and D. Harneldi, "Experimental study on improved parametric stereo for bit rate scalable audio coding," in *Proc. of 2014 Int. Conf. on Information Tech. and Electrical Eng.*, Jogjakarta, Indonesia, October 2014.

[17] D. P. Chen, H. F. Hsiao, H. W. Hsu, and C. M. Liu, "Gram-schmidt-based downmixer and decorrelator in the MPEG surround coding," Presented at the 128^{th} Convention of the Audio Engineering Society, London, UK, May 2010.

[18] J. Hilpert and S. Disch, "The MPEG Surround audio coding standard [Standards in a nutshell]," *IEEE Signal Processing Mag.*, vol. 26, no. 1, pp. 148–152, Jan. 2009.

[19] J. Herre *et al.*, "MPEG Surround - The ISO/MPEG standard for efficient and compatible multichannel audio coding," *J. Audio Eng. Soc.*, vol. 56, no. 11, pp. 932–955, 2008.

[20] I. Elfitri, M. Muharam, and M. Shobirin, "Distortion analysis of hierarchical mixing technique on MPEG surround standard," in *Proc. of 2014 Int. Conf. on Advanced Computer Sciences and Information System*, Jakarta, Indonesia, October 2014.

[21] M. Wolters, K. Kjorling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," in *Proc. the* 115^{th} *Convention of the Audio Engineering Society*, New York, USA, October 2003.

[22] J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding," *IEEE Signal Proc. Mag.*, vol. 25, no. 3, pp. 137–142, 2008.

[23] J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjorling, J. Plogsties, and J. Koppens, "Multi-channel goes mobile: MPEG Surround binaural rendering," in *Proc. the AES* 29^{th} *Int. Conference*, Seoul, Korea, September 2006.

[24] J. Herre and L. Terentiv, "Parametric coding of audio objects: Technology, performance, and opportunities," Presented at the 42^{nd} Int. Conference: Semantic Audio, Ilmenau, Germany, July 2011.

[25] J. Herre, C. Falch, D. Mahne, G. del Galdo, M. Kallinger, and O. Thiergart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," Presented at the 128^{th} Convention of the Audio Engineering Society, London, UK, May 2010.

[26] G. Hotho, L. F. Villemoes, and J. Breebaart, "A backward-compatible multichannel audio codec," *IEEE Trans. Audio, Speech Language Process.*, vol. 16, no. 1, pp. 83–93, Jan. 2008.

[27] J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen, and S. V. de Par, "Background, concepts, and architecture for the recent MPEG Surround standard on multichannel audio compression," *J. Audio Eng. Soc.*, vol. 55, pp. 331–351, 2007.

[28] ITU-R, "Method for Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," Recommendation ITU-R BS.1116-1, 1997.