ISBN: 978-1-4799-5302-8 Yogyakarta, 7-8 October 2014 ICITEE 2014

Investigation on Objective Performance of Closed-loop Spatial Audio Coding

Ikhwana Elfitri, Rahmadi Kurnia, and Fitrilina

Department of Electrical Engineering, Andalas University
Kampus Limau Manis, Padang, 25163, Indonesia
Email: ikhwana@ft.unand.ac.id

***Abstract*—Closed-loop spatial audio coding is a compression technique, developed based on MPEG Surround (MPS) standard, having an advantage of minimising distortion due to quantisation process of spatial parameters. Despite the MPS is developed based on filterbank, however, this closed-loop system performs better on Modified Discrete Cosine Transform (MDCT). Considering its high performance over the open-loop system, this paper presents further investigation on objective performance of closed-loop spatial audio coding against various quantisers of spatial parameters. Experiments have been conducted to measure signal to noise ratio (SNR) across different types of uniform spatial quantisers at various operating bitrates. The results show that the SNR achieved by the open-loop approach is strongly affected by the type of the quantiser while, in contrast, the SNR achieved by the closed-loop approach is relatively constant regardless the number of bits used in the quantisers. Moreover, the results also show that the closed-loop configuration can consistently improve SNR in any quantisation scheme.**

***Keywords*—Spatial audio coding, MPEG Surround, Closed-loop system.**

I. Introduction

Considering the growing demand for the reliable delivery of high quality multichannel audio in various multimedia applications such as home entertainment, digital audio broadcasting, computer games, music streaming services as well as teleconferencing, efficient coding techniques [1]–[3] have become essential for advanced audio processing. The traditional approach for compressing multichannel audio is to encode each audio channel using a mono audio coder, such as Dolby AC-3 [4] and MPEG advanced audio coder (AAC) [5], [6]. However, for the majority of coders adopting this method, the number of bits to be transmitted tends to increase linearly with the number of channels.

Recently, a new concept for encoding multichannel audio signals has been proposed. It comprises the extraction of the spatial cues and the downmixing of multiple audio channels into a mono or stereo audio signal [7]. The downmix signals are subsequently compressed by an existing audio encoder and then transmitted, accompanied by the spatial cues coded as spatial parameters. Any receiver system that cannot handle multichannel audio can simply remove this side information and just render the downmix signals. This provides the coder with backward compatibility, which is important for implementation in various legacy systems. In addition, by utilising the spatial parameters, the downmix signals can be directly upmixed at the decoder side into a multichannel configuration that may be different from the one used at the encoder side. This technique is known as spatial audio coding (SAC) [8], [9].

Various SAC techniques, such as binaural cue coding, (BCC) [10], [11] and MP3 Surround [12], [13] have been proposed. Inter-channel level difference (ICLD), inter-channel time difference (ICTD), and inter-channel coherence (ICC) are extracted as spatial parameters that are based on human spatial hearing cues [14]. Techniques such as parametric stereo (PS) [15] and MPEG Surround (MPS) [16], [17] may also utilise signal processing techniques, such as decorrelation. The great benefit of these perceptual-based coders is that they can achieve bitrates as low as 3 kb/s for transmitting spatial parameters, as in the case of MPS [16].

Based on MPEG Surround, a closed-loop spatial audio coding technique [18], [19] has been proposed and its performance, tested objectively and subjectively, has been reported to show the advantages of the closed-loop system in order to improve the quality of multichannel audio reproduction. However, it is not clear yet how the effect of different spatial parameter quantisers on the performance of the closed-loop codec. In this paper, a further investigation on the objective performance of the closed-loop spatial audio coding against various quantisers of spatial parameter is presented to provide a comprehensive study on its performance. This paper is organised as follows. In Section II an overview of the closed-loop system is briefly discussed foolowed in Section III by a brief introduction on its performance. In Section IV the results of the experiments are presented while the conclusion is given in Section V.

II. Overview of the Closed-loop Spatial Audio Coding

Fig. 1 shows a block diagram of an L-channel audio encoder applying a closed-loop configuration. As applied in the MPEG Surround, this closed-loop system comprises a pair of elementary building blocks for channel conversion and the reverse process: one-to-two (OTT) and reverse OTT (R-OTT). The OTT module is used to convert a single channel into two channels while the reverse conversion is done by the R-OTT module. For encoding more than 2 channels, the whole process is undertaken by combining a number of OTT in a tree structure. This section discusses the process of extraction of the CLD, ICC as well as the residual signal. It is necessary to notice that the modified discrete cosine transform (MDCT) is proposed to replace the hybrid filterbank to achieve

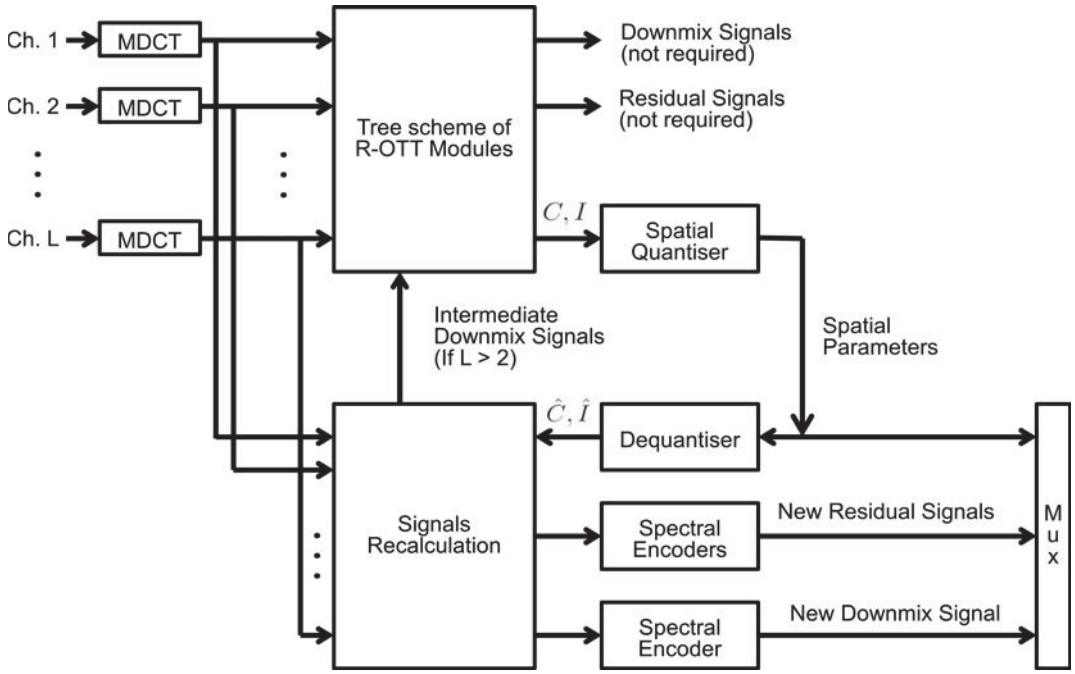

Fig. 1. The block diagram of L-channel MDCT-based closed-loop encoder.

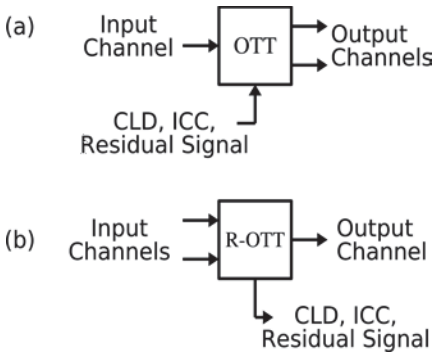

Fig. 2. Block diagram of (a) OTT module and (b) R-OTT module as used in MPS.

higher performance as it will be shown in the results of the experiments. For this MDCT-based implementation, a number of scale factor bands as defined in [20] is applied.

Both R-OTT and OTT modules are shown in Fig. 2. On one hand, the R-OTT module is applied in the encoder side to extract spatial parameters: Channel level difference (CLD), Interchannel coherence (ICC). Then, the downmix signal is generated. For compensating error due to downmixing process, residual signal can be transmitted. On the other hand, the OTT is applied in the decoder side to recreate two audio channels from the single downmixed channel based on the received spatial parameters.

The CLD, a comparison of signal energy in the first channel, $x_1[k]$, to signal energy in the second channel, $x_2[k]$, can be calculated for a parameter band s, as follow:

$$C(s) = \frac{e^s_{x_1}}{e^s_{x_2}} = \frac{\sum\limits_k x_1(s,k) \cdot x_1^*(s,k)}{\sum\limits_k x2(s,k) \cdot x_2^*(s,k)} \tag{1}$$

where k is the index of the spectral coefficients as output of the MDCT transformation and the sign of $(^*)$ represents a complex conjugate. The energy ratio is then represented in logarithmic for the purpose of parameter transmission.

The ICC parameter that is a correlation of both input channels, indicated with I variable, can be calculated as the real number of the following equation:

$$I(s) = \frac{\sum_k x_1(s,k) \cdot x_2^*(s,k)}{\sqrt{e^s_{x_1} e^s_{x_2}}} \tag{2}$$

Moreover, the downmix signal is a sum of both input channels where each channel is individually weighted. The closed-loop representation [19] of the downmix signal $y[n]$, can be written as:

$$y(k) = \frac{x_1(k) + x_2(k)}{\hat{a}^s + \hat{b}^s} \tag{3}$$

For this downmixing process, the energy constants a and b [21] are introduced to preserve the energy constraint and can be calculated as follow:

$$(a^s + b^s)^2 = \frac{e^s_{x_1} + e^s_{x_2} + 2I\sqrt{e^s_{x_1} e^s_{x_2}}}{e^s_{x_1} + e^s_{x_2}} \tag{4}$$

Furthermore, the residual signal $r[n]$ in each subband is determined from the following decomposition:

$$x_1^s[n] = a^s y^s[n] + r^s[n] \tag{5a}$$
$$x_2^s[n] = b^s y^s[n] - r^s[n] \tag{5b}$$

TABLE I. MAPPING OF 49 SCALE FACTOR BANDS TO 20 PARAMETER BANDS

PB Index	SFB Index	PB Index	SFB Index
1	1	11	16-17
2	2	12	18-19
3	3	13	20-21
4	4	14	22-23
5	5-6	15	24-25
6	7-8	16	26-27
7	9-10	17	28-29
8	11	18	30-31
9	12-13	19	32-37
10	14-15	20	38-49

which produces a single residual signal for reconstructing both $x_1^s[n]$ and $x_2^s[n]$.

At the decoder side, both audio signals are recreated by estimating a and b as follows:

$$\hat{a} = X\cos(A+B) \tag{6a}$$

$$\hat{b} = Y\cos(A-B) \tag{6b}$$

where the X, Y, A, and B variables given as

$$X = \sqrt{\frac{\hat{C}}{1+\hat{C}}} \tag{7a}$$

$$Y = \sqrt{\frac{1}{1+\hat{C}}} \tag{7b}$$

$$A = \frac{1}{2}\arccos(\hat{I}) \tag{7c}$$

$$B = \tan\left[-\left(\frac{X-Y}{X+Y}\right)\arctan(A)\right] \tag{7d}$$

are determined from the quantised values of CLD, $\hat{C}$, and the quantised values of ICC, $\hat{I}$. Hence, both signals can be reconstructed as

$$\hat{x}_1[n] = \hat{a}\hat{y}[n] + \hat{r}[n] \tag{8a}$$

$$\hat{x}_2[n] = \hat{b}\hat{y}[n] - \hat{r}[n] \tag{8b}$$

which are similar to (5) but use the decoded downmix and residual signals, $\hat{y}[n]$ and $\hat{r}[n]$, respectively. The indices for the subbands and parameter bands have been ignored for notation simplicity.

In order to provide a scalable spatial resolution, the MDCT-based closed-loop R-OTT module can be applied using a number of parameter bands which is proposed, for simplicity, to be equal to the number of parameter bands of the MPS. For instance, such a mapping for grouping 49 Scale Factor Bands (SFB) to 20 Parameter Bands (PB) is given in Table I.

III. OBJECTIVE PERFORMANCE OF THE CLOSED-LOOP METHODE AT HIGH BITRATE

TABLE II. AVERAGE SEGSNR (DB) OF VARIOUS R-OTT METHODS

Audio Channel	FB-Based OL R-OTT	FB-Based CL R-OTT	MDCT-Based CL R-OTT
Left	28.83	31.55	38.48
Right	24.21	29.50	36.64
Average	26.52	30.52	37.56

TABLE III. AVERAGE SEGSNRS (IN DECIBELS) OF MDCT-BASED CODECS COMPARED TO FB-BASED R-OTT

Audio Channel	FB-Based OL R-OTT	FB-Based CL R-OTT	MDCT-Based CL R-OTT
Left	26.38	32.06	38.00
Right	25.49	31.27	38.02
Centre	20.64	28.63	36.65
Left surround	22.25	31.98	38.15
Right surround	26.92	31.70	38.27
Mean	24.34	31.13	37.82
STD	2.75	1.43	0.66

Table II demonstrates that the proposed closed-loop R-OTT module is capable of improving the segSNR when operating at 160 kb/s per audio channel. It shows that the FB-based closed-loop R-OTT can improve the segSNR by approximately 4 dB. Moreover, performing the MDCT-based closed-loop R-OTT module is able to improve further the segSNR. About 11 dB of segSNR improvement is achieved in comparison to the FB-based open-loop R-OTT. Moreover, it is shown in Table III that the MDCT-based closed-loop R-OTT is able to reach the average segSNR of 37.82 dB. Approximately 13 dB of SNR improvement is obtained, compared to the conventional FB-based open-loop R-OTT module. Benchmarking to the FB-based closed-loop R-OTT module shows that the segSNR increases by more than 6 dB. The results indicate that the closed-loop R-OTT algorithm is also capable of minimising signal distortion at all channels of 5-channel signals resulting in segSNR improvement.

Interesting results were produced in situations where the segSNR achieved by the MDCT-based closed-loop R-OTT in each channel has similar values. The average segSNR is 37.82 dB, while the standard deviation is equal to 0.66 dB. This standard deviation is significantly lower than 2.75 dB which is the standard deviation of the segSNR measured on the FB-based open-loop R-OTT module. It is also considerably lower than 1.43 dB: the standard deviation of the segSNR achieved by the FB-based closed-loop R-OTT module. The results indicate that the encoder performing the MDCT-based closed-loop R-OTT module does not only significantly improve the segSNR but also makes the segSNR of audio signals very close to each other. It suggests that the coder is able to distribute the distortion uniformly across the channels.

IV. RESULTS

A. Experimental Setup

Experiments have been conducted to investigate the effect of various quantisers in quantising spatial parameters on the performance of the closed-loop tehcnique. An audio excerpt (high-correlated audio signals) was used as inputs for the experiments. This audio excerpt consists of the Left (L), Right (R), Centre (C), Left surround (Ls), and Right surround (Rs) channels of 5.1 recordings containing panned mixtures of the individual audio objects. For 2-channel input, the left (L) and right (R) channels were used as inputs. In calculating CLDs and ICCs, 20 parameter bands were used in both subband and frequency domain implementation.

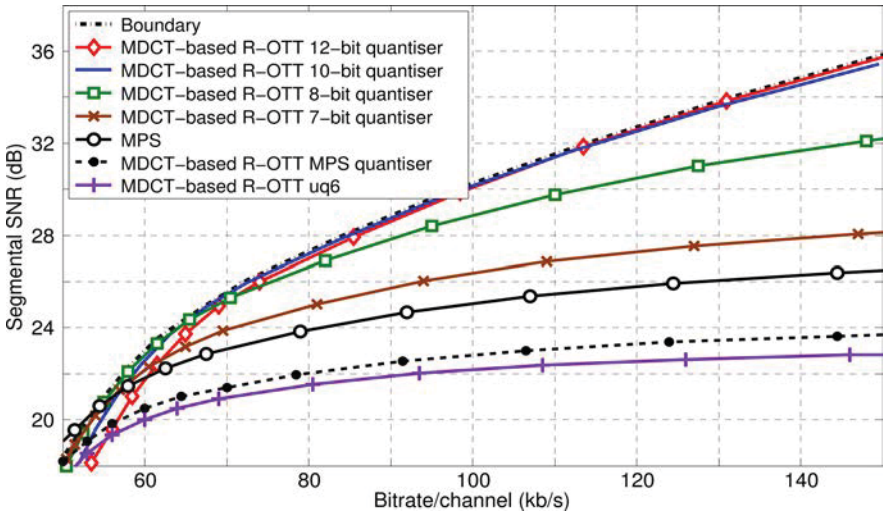

Fig. 3. SegSNR of open-loop MDCT-based R-OTT module for various quantisers.

B. Evaluation of the MDCT-based Open-loop R-OTT Module

The experiments in this section were conducted by operating a number of audio codecs, employing the MDCT-based open-loop R-OTT module, using different quantisers. All audio codecs encoding 2-channel audio signals operated at various bitrates ranging from 50 to 150 kb/s per audio channel. The CLDs and ICCs were quantised uniformly. As many as 5 different quantisation tables, each with its own stepsize, were tested. For benchmarking, an audio codec employing the subband domain open-loop R-OTT module (i.e. as implemented in MPS), operating at the same range of bitrates, was also tested. In addition the MDCT-based R-OTT module was also tested using the MPS's spatial quantiser. The average segSNRs measured on the experiments are plotted in Fig. 3.

The results show that the audio codecs applying the MDCT-based open-loop R-OTT module quantising and representing CLD and ICC with 7, 8, 10 and 12 bits achieve higher segSNRs than the audio codec employing the subband domain open-loop R-OTT module at almost all operating bitrates. On the other hand, the audio codecs applying the MDCT-based open-loop R-OTT module, quantising each CLD and ICC using 6-bit uniform quantiser, as well as the MPS' spatial quantiser, have lower segSNRs than the 2-channel MPS at all operating bitrates. The results clearly indicate that the implementation of the MDCT-based open-loop R-OTT module at the tested bitrates can achieve a higher performance than the MPS, provided that appropriate spatial quantisers are applied. It suggests that the spatial quantisers play an important role in enhancing the performance of the audio codec applying the open-loop R-OTT module.

Furthermore, it is interesting to note that the implementation of the MDCT-based R-OTT module employing MPS's spatial quantisers cannot achieve better performance than the MPS itself. It clearly indicates that the MPS's spatial quantisers are not appropriate for the implementation of the MDCT-based open-loop R-OTT module. It suggests that for applying the R-OTT module in the frequency domain, an investigation should be conducted for implementation over various spatial quantisers. Then proper spatial quantisers should be chosen for a particular range of operating bitrates in order to achieve the optimal performance of the MDCT-based R-OTT module.

It can also be seen that at a bitrate of 150 kb/s per audio channel, approximately 9 dB of SNR improvement, compared with MPS, is achieved by the codec employing the MDCT-based R-OTT module using 12-bit uniform spatial quantiser. Conversely, a slightly different performance is shown by the audio codec when applying a 10-bit uniform spatial quantiser. However, the segSNR improvement decreases as the bitrates become lower. It also shows that more than 3 dB segSNR improvement is achieved by the audio codecs using both 10 and 12-bit uniform spatial quantisers operating at bitrates between 80 and 150 kb/s per audio channel. Furthermore, the segSNR improvement continues to decrease and becomes insignificant (i.e. improvement is around 1 dB) at lower bitrates, particularly at around 65 kb/s per audio channel. In contrast, at operating bitrates lower than 60 kb/s per audio channel, the MPS achieves the highest segSNR among all tested audio codecs.

The results indicate that the MDCT-based open-loop R-OTT module is not appropriate for low bitrate implementation below 60 kb/s per audio channel. It suggests that the performance of the R-OTT module applied in the frequency domain is subject to the operating bitrate. Moreover, it also suggests that the open-loop R-OTT module performs better in the subband domain than in the frequency domain, when the bitrate is lower than 60 kb/s per audio channel.

In addition, Fig. 3 also shows a boundary of the segNSRs that can be achieved by the audio codec applying any number of bits applied in the uniform spatial quantiser to represent the CLD and ICC. This boundary may be useful to grade the performance of the proposed MDCT-based closed-loop R-OTT module.

C. Comparison of the Open and Closed-loop R-OTT Modules over Various Spatial Quantisers

Experiments in this section were conducted by measuring the segSNR for different quantisation errors. The audio codecs employing the MDCT-based closed-loop R-OTT module using

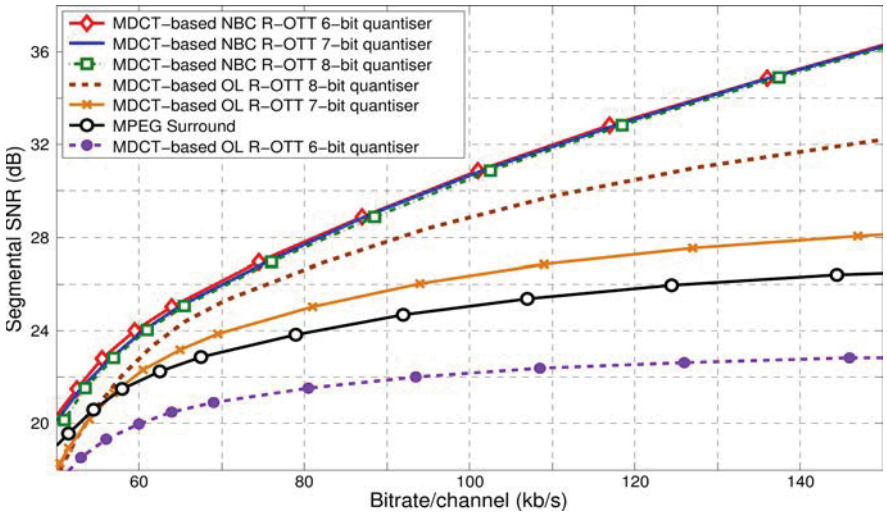

Fig. 4. Comparison of open-loop and closed-loop R-OTT modules over 3 types of spatial quantisers.

3 different uniform spatial quantisers representing both the CLDs and ICCs with 6, 7, and 8 bits were tested. For comparison, the codecs applying the MDCT-based open-loop R-OTT using similar spatial quantisers were also included in the experiments. The result were plotted and depicted in Fig. 4. It shows that the closed-loop technique significantly improves the average segSNR for every type of the tested quantisers. As expected, the results demonstrate that the performances of the closed-loop R-OTT modules using different spatial quantisers are similar. As previously explained, the closed-loop R-OTT module operates by minimising the errors contributed during the quantisation of the spatial parameters. Hence, the stepsize of the spatial quantiser, which leads to the amount of the quantisation error, does not affect the performance of the closed-loop R-OTT module. It indicates that the closed-loop R-OTT module is capable of greatly minimise the error introduced by the spatial quantiser regardless of the amount of error introduced. It therefore suggests that the best spatial quantiser, in terms of the smallest number of bits required, among the tested quantisers to be applied in the MDCT-based R-OTT module, is the MPS's spatial quantisers.

D. Evaluation of the Closed-loop R-OTT Module Using the MPS's Spatial Quantisers

In order to investigate further the performance of the MDCT-based closed-loop R-OTT module using the MPS's spatial quantisers, experiments have also been conducted for encoding 2-channel audio signals over various operating bitrates between 50 and 150 kb/s per audio channel. The results of the experiments are given in Fig. 5. For comparison, the upper bounds of the segSNRs achieved by the MDCT-based open-loop R-OTT module, termed as boundary, are plotted. Moreover, the segSNRs of the FB-based open and closed-loop R-OTT modules are also included. The results demonstrate that the closed-loop R-OTT module can improve segSNR in the subband and even further in the frequency domain. Moreover, it is also shown that the MDCT-based closed-loop R-OTT module can exceed the upper bound of the segSNRs that can be achieved by the MDCT-based open-loop R-OTT module. It indicates that the closed-loop R-OTT module applying the MPS's quantisers consistently has a higher segSNR than the open-loop R-OTT module, regardless of the spatial quantiser applied.

V. Conclusion

This paper has presented the results of experiments to investigate objective performance of the closed-loop MDCT-based spatial audio coding across differents spatial parameter quantisers. Segmental signal to noise ratio (segSNR) was used as an objective metric. The maximum segSNR that can be achieved by the open-loop methode has been presented. Moreover, the amount of segSNR improvement that achieved by applying the closed-loop system has also been shown for different quantisers. It has also been demonstrated that the spatial quantiser, that is implemented in MPEG Surround, is the most suitable one for the MDCT-based spatial audio coding technique.

Acknowledgment

This work was sponsored in part by the Ministry of Education and Culture, Republic of Indonesia, under DIPA Universitas Andalas, contract no. 023.04.2.415061/2014 while some experiments have been conducted in the I-Lab Multimedia Communication Research, the University of Surrey, Guildford, GU2 7XH, UK.

References

[1] 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.

[2] M. Bosi, "High-quality multichannel audio coding: trends and challenges," *J. Audio Eng. Soc.*, vol. 48 No. 6, pp. 588–595, 2000.

[3] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–513, April 2000.

[4] C. C. Todd, G. A. Davidson, M. F. Davis, L. D. Fielder, B. D. Link, and S. Vernon, "AC-3: Flexible perceptual coding for audio transmission and storage," in *Proc. the* 96^{th} *Convention of the Audio Engineering Society*, Amsterdam, The Netherlands, 1994.

[5] 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.

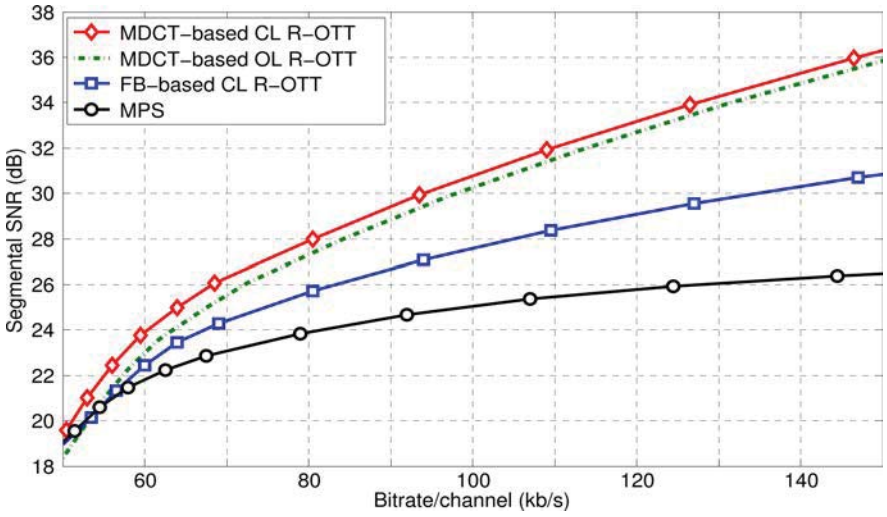

Fig. 5. SegSNR of MDCT-based closed-loop R-OTT module using MPS's spatial quantisers compared to open-loop systems.

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

[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] F. Baumgarte and C. Faller, "Binaural cue coding-part I : Psychoacoustic fundamentals and design principles," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 509–519, Nov. 2003.

[11] C. Faller and F. Baumgarte, "Binaural cue coding-Part II : Schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.

[12] B. Grill, O. Hellmuth, J. Hilpert, J. Herre, and J. Plogsties, "Closing the gap between the multichannel and the stereo audio world: Recent mp3 surround extensions," in *Proc. the* 120^{th} *Convention of the Audio Engineering Society*, Paris, France, May 2006.

[13] H. Moon, "A low-complexity design for an mp3 multichannel audio decoding system," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 20, no. 1, pp. 314–321, January 2012.

[14] J. Blauert, *Spatial Hearing, The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 2001.

[15] 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.

[16] J. Herre, K. Kjorlings, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Linzmeier, and K. S. Chong, "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.

[17] 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.

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

[19] 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.

[20] ISO/IEC, "Information Technology - Coding of audio-visual objects, Part 3: Audio," ISO/IEC 14496-3:2009(E), International Standards Organization, Geneva, Switzerland, 2009.

[21] 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.