Experimental Study on Improved Parametric Stereo for Bit Rate Scalable Audio Coding

Ikhwana Elfitri, Rahmadi Kurnia, and Defry Harneldi

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

***Abstract*—This paper presents an experimental study on generalizing Parametric Stereo (PS) technique in an attempt to make it scalable for low and high bit rate implementations which are very useful for various internet services. As specified in the MPEG standard, PS is only intended for applications at a stereo bit rate of 24 kb/s. If operated at higher bit rates, PS would have much worse performance than other audio codecs such as MPEG-1/2 Layer 3 (MP3) and MPEG Advanced Audio Coding (AAC). To improve performance at higher bit rates, the proposed Improved PS transmits residual signal. However, the bit rates of the down-mix and residual signals have to be selected properly to achieve optimal performance. The results of the experiments show that performance of the Improved PS, in terms of objective difference grade (ODG), significantly improves as the bit rate increases. Testing objectively using 5 different critical audio materials, Improved PS consistently outperforms AAC Stereo at a bit rate of 128 kb/s.**

Keywords*—*Audio Compression, Parametric stereo, MPEG advanced audio coding (AAC).

I. INTRODUCTION

Perceptual audio coding [1]–[3] is essential for various internet and communication applications such as music downloading services, teleconference, audio streaming, and digital audio broadcasting. The purpose is to represent and deliver audio signal as a few bits as possible. Even though the speed of the internet services tend to increase, low bit rate audio codec is still favourable. However, it is often desirable that the audio codec can be scalable in terms of operating bit rate, making it possible to operate at both low and high bit rate implementation.

Amid the increasing popularity of spatial audio [4]–[6] where more than 2 channels of audio signals are normally employed, the representation of audio scene as 2 channels or stereo is still widely used due to its simplicity in that internet users can render the audio easily through their mobile device accompanied with a headset. A number of stereo audio coding standards are currently available such as MPEG 1/2 Layer 3 (MP3) [7] and MPEG Advanced Audio Coding (AAC) [8], [9]. The MP3, as the first standard for audio coding, is the most popular even though the quality of the MP3 compressed audio signal is significantly lower than the AAC compressed audio signal particularly for operating at low bit rates i.e at bit rates of 64 kb/s/channel and lower. The MPEG AAC itself consists of 3 variants: Low Complexity (LC) AAC, High Efficiency (HE) AAC [10], which utilises Spectral Band Replication (SBR), and HE-AAC v2 which applies Parametric Stereo (PS) technique [11]–[14].

All of those AAC variants are typically intended for different operating bit rates. The LC-AAC was reported as capable of reproducing indistinguishable audio signal at a bit rate of 64 kb/s/channel. On one hand, the HE-AAC generally performs better than LC-AAC at various bit rates lower than 64 kb/s/channel. On the other hand, the HE-AAC v2 or Parametric Stereo was intended to operate at a bit rate of 12 kb/s/channel or at a total bit rate of 24 kb/s for both stereo channels. Depending on the characteristic of the input audio signal, the quality of the audio signal reconstructed by each variant may differ at its typical bit rate. However, none of them can achieve the highest performance at all operating bit rates.

This paper presents an investigation on the possibility of improving Parametric Stereo technique to operate at higher bit rates. It will be shown that the Parametric Stereo technique can perform better than LC-AAC at various bit rates that are much higher than its typical operating bit rate. This investigation also reveals that different bit allocation is required for each tested audio material. The capability of PS technique to achieve high performance at high bit rate implementation provides bit rate scalability that is an ability to operate a single audio coding method for various operating bit rates.

II. OVERVIEW OF PARAMETRIC STEREO (PS)

PS technique can be briefly explained as a method to downmix a stereo audio signal into a mono audio signal while stereo parameters are extracted in order to be able to recreate the stereo signal at the decoder side. Subsequently, the mono audio signal can be generally encoded by any type of audio encoder, however, in the MPEG standard, HE-AAC is employed. Then, the mono downmixed signal is transmitted along with stereo parameters. At the decoder side, the mono audio signal and stereo parameters are first decoded. The stereo audio signal can be reconstructed by upmixing the mono audio signal utilising the stereo parameters. Decorrelator is employed to create a synthetic signal that can minimize the distortion due to downmixing and upmixing processes.

A. PS Encoder

Fig. 1 shows a block diagram of PS technique. Two audio channels in stereo format are first decomposed by analysis filter bank resulting in 71 subband audio signals. For the purpose of parameter extraction, parameter band is defined as

Fig. 1. Block diagram of Parametric Stereo: encoder (left) and decoder (right)

a group of subband signals. The number of subband signals for each parameter band can be different each other which is actually determined based on the critical bandwidth. A total of either 10, 20, or 34 parameter bands can be applied. For each parameter band, 4 stereo parameters are extracted: interchannel intensity difference (IID), interchannel coherence (ICC), interchannel phase difference (IPD) and overall phase difference (OPD).

For stereo parameter estimation in a parameter band b, the following expressions are calculated:

$$e_l(b) = \sum_k \sum_n [l(k,n)l^*(k,n)] + \epsilon \tag{1a}$$

$$e_r(b) = \sum_k \sum_n [r(k,n)r^*(k,n)] + \epsilon \tag{1b}$$

$$e_R(b) = \sum_k \sum_n [l(k,n)r^*(k,n)] + \epsilon \tag{1c}$$

$$e_O(b) = \sum_k \sum_n [l(k,n)m^*(k,n)] + \epsilon \tag{1d}$$

where $l(k,n), r(k,n), m(k,n) = [l(k,n), r(k,n)]/2$ are subband signals in the left, right, and mono output channels respectively, k is subband index, n is the sample index of subband signal, the * is a symbol for complex conjugate, and $\epsilon = 1e^{-10}$ is a constant added to avoid the case of divided by zero in the computation.

Based on those expressions, stereo parameters can be calculated. Interchannel intensity difference for a parameter band b, $iid(b)$, can be calculated as:

$$iid(b) = \frac{e_l(b)}{e_r(b)} \tag{2}$$

Next, interchannel coherence for a parameter band b, $icc(b)$, is determined by:

$$icc(b) = \sqrt{\frac{[e_R(b)]^2}{e_l(b)e_r(b)}} \tag{3}$$

Furthermore, interchannel phase difference for a parameter band b, $ipd(b)$, is estimated as below:

$$ipd(b) = \angle e_R(b) \tag{4}$$

where $\angle$ is a symbol for four quadrant angle. And, overall phase difference for the parameter band b, $opd(b)$, is obtained by:

$$opd(b) = \angle e_O(b) \tag{5}$$

B. PS Decoder

Two channels stereo subband signals, the left, $l(k,n)$, and the right, $r(k,n)$, can be constructed from the decoded downmix signal, $s(k,n)$, and the decorrelated signal, $d(k,n)$, using the following 4 matrices: $H_{11}, H_{12}, H_{21}, H_{22}$, as below:

$$l(k,n) = H_{11}(b)s(k,n) + H_{21}(b)d(k,n) \tag{6a}$$

$$r(k,n) = H_{12}(b)s(k,n) + H_{22}(b)d(k,n) \tag{6b}$$

where these matrices are determined from the decoded stereo parameters as detailed below:

$$H_{11}(b) = \sqrt{2}\cos(\alpha(b))\cos(\beta(b))\exp(j \cdot opd(b)) \tag{7a}$$

$$H_{12}(b) = \sqrt{2}\sin(\alpha(b))\cos(\beta(b))\exp(j \cdot (opd(b) - ipd(b))) \tag{7b}$$

$$H_{21}(b) = \sqrt{2}\sin(\alpha(b))\sin(\beta(b))\exp(j \cdot opd(b)) \tag{7c}$$

$$H_{22}(b) = \sqrt{2}\cos(\alpha(b))\sin(\beta(b))\exp(j \cdot (opd(b) - ipd(b))) \tag{7d}$$

with α and β are found from ICC and IID as below:

$$\alpha(b) = \frac{1}{2}\arccos(icc(b)) \tag{8}$$

$$\beta(b) = \alpha(b)\frac{1 - iid(b)}{\sqrt{1 + [iid(b)]^2}} \tag{9}$$

The IPD and OPD parameters are actually optional. If the encoder does not transmit these phase parameters then the ipd and opd will be set to zero. Otherwise, they have to be smoothed over time [8] which is not further discussed here for simplicity. The reconstructed subband stereo signals are then transformed back to fullband time domain audio signal by synthesis filter bank.

III. Improved Parametric Stereo

Fig. 2 shows a block diagram of the Improved PS by enabling the residual signal transmission. This approach has been adopted from the one-to-two (OTT) module of MPEG Surround [15]–[17]. The main purpose of the Improved Parametric Stereo is to transmit residual signal for compensating the error due to downmixing stereo signal to mono signal. Depending on the operating bit rate, fullband or lower bandwidth of residual signal can be transmitted to decoder. In order to achieve the optimal combination of downmix and residual signals, a bit allocator is proposed to be applied in the mono processing block. However, it is currently being further investigated and more details on the methode employed in the bit allocator is out of the scope of this paper.

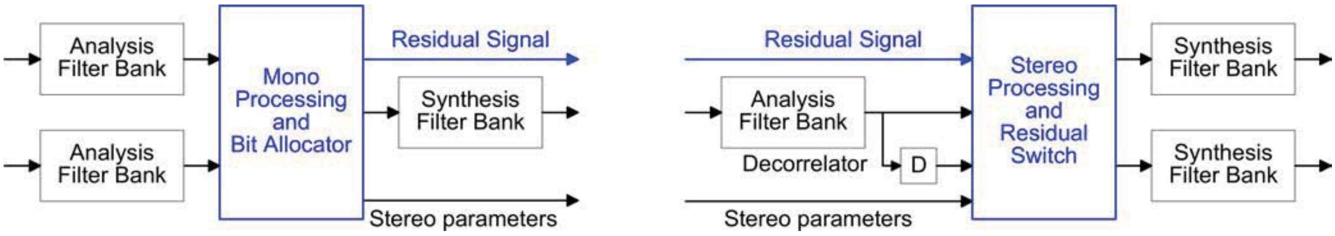

Fig. 2. Block diagram of Improved Parametric Stereo: encoder (left) and decoder (right). The blue boxes are parts of standard that have been modified for using residual signal.

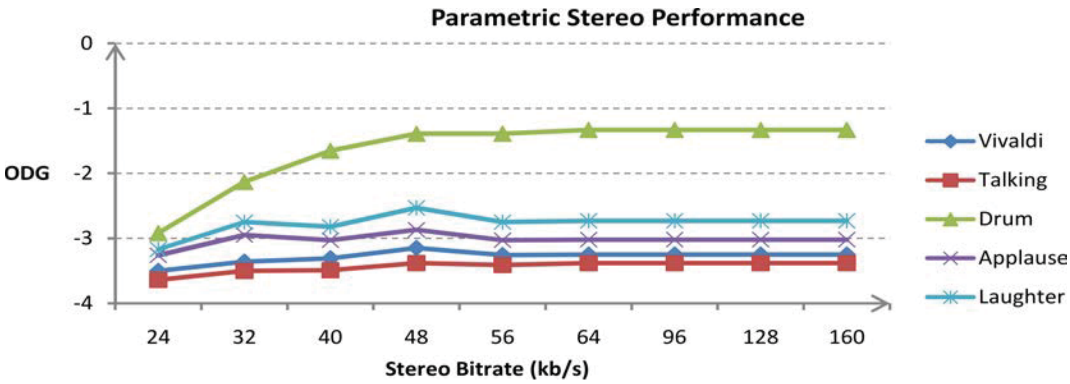

Fig. 3. Parametric Stereo performance at various bit rates.

TABLE I. List of Audio Excerpts for Experiments

Excerpt Name	Description
Applause	Hundreds people clap their hand
Drum	Acoustic music: drum, guitar, male vocal
Laughter	Sound of hundreds of people laughing
Talking	Male and female speeches + music background
Vivaldi	Classical music with vocal

At the decoder side, a mechanism called residual switch is applied in the stereo processing block, to choose and switch between the residual signal and the decorrelated signal. The role of this switch is to search the existence of the residual signal first. In a condition when residual signal is not received then the decorrelated signal will be employed. At the encoder side, the subband residual signal, $d(k, n)$, is determined based on the following decomposition:

$$l(k, n) = \varepsilon_1(b)m(k, n) + d(k, n) \tag{10a}$$

$$r(k, n) = \varepsilon_2(b)m(k, n) - d(k, n) \tag{10b}$$

where $\varepsilon_1(b)$ and $\varepsilon_2(b)$ are energy constants that depend on the IID. For this Improved PS scheme, the IPD and OPD are disabled for simplicity.

IV. Experimental Results

A number of experiments have been conducted to investigate the performance of the existing PS standard and the performance of the proposed Improved PS. The AAC stereo was used as a benchmark. The PS and the AAC as implemented in Nero audio codec was employed in the experiments. Five critical audio materials as listed in Table I were used. The performance is assessed in terms of objective difference grade (ODG) [18] score. As many as 20 parameter bands were used while the stereo parameters were transmitted at a bit rate of 8 kb/s.

A. Analysis of Existing PS Standard

In order to show the performance of existing PS standard at high bit rates, experiments have been conducted to assess its performance at bit rates ranging from 24 to 160 kb/s. The results are given in Fig. 3. It can be seen that the PS standard cannot achieve ODG score higher than -1 (excellent). Other than the drum audio excerpt, the PS standard achieves even lower scores i.e. less than -2 (good). The results also show that at bit rates of 64 kb/s and higher, increasing the bit rate cannot improve ODG score. It suggests that the PS standard is not appropriate for high bit rate implementation.

B. Comparison of Improved PS and AAC

The experiments in this section are intended to investigate the performance of Improved PS at a bit rate of 128 kb/s. To provide initial results, a number combinations of bit rate for downmix and residual signals were selected experimentally and then tested. Five combinations that provide the highest ODG score, are presented in Fig. 4. These bit rate combinations are shown in the horizontal axis. For example, 64-56 shows that the downmix signal was transmitted at a bit rate of 64 kb/s while the residual signal was transmitted at a bit rate of 56 kb/s. The results also present the performance of AAC Stereo, operated at 128 kb/s without bit rate combinations, for benchmarking.

The results show: first, the proposed Improved PS codec can achieve ODG scores closer to zero suggesting that it can operate at higher bit rates with excellence performance. It suggests that PS can be used as a bit rate scalable audio codec operating for a wide range bit rates. Second, PS codec performance differs for every combination of bit rates for downmix and residual signal. The highest ODG score for each tested audio material is also different. The results also show that the ODG scores tend to be high when the downmixed signal is allocated more bits than the residual signal. For instance, the downmix signal can be transmitted at a bit rate as high as 104 kb/s. However, for some combinations, the downmix signal can also be transmitted at a bit rate lower than the bit rate of the residual signal such as at a bit rate of 40 kb/s, as shown when tested with the talking audio excerpt. Third, the performance of PS and AAC codecs differs for each tested audio materials. However, it can be seen that for every tested audio material, PS codec achieves better ODG score than AAC. When tested using the vivaldi excerpt, PS codec can achieve a slightly higher score than AAC. However, the results show that testing PS codec with the drum and laughter audio excerpts demonstrates significant ODG score improvement where more than 0.5 ODG score is achieved by PS codec compared to AAC.

In general, Improved PS achieves higher ODG scores than AAC for every tested audio excerpt. However, the finding for the optimal combination of bit allocation for the downmix and residual signals seems to be important in order to achieve the highest ODG score. Otherwise, the ODG score achieved by the Improved PS technique can be lower than AAC as it is shown in the results when the vivaldi audio excerpt was used as input signal. It can be seen that at a combination of 96-24 kb/s, AAC codec has higher ODG score than Improved PS codec. The results for this combination also show that increasing the bit rate of the downmix signal cannot ensure of scoring the highest ODG.

V. Conclusions

An Improved Parametric Stereo technique has been presented in this paper. This approach was designed to facilitate Parametric Stereo (PS) as a universal coding algorithm for all operating bit rates which are necessary for various internet services. In order to achieve this aim, the residual signal is transmitted and bit allocator is rrequired to choose an optimal bit allocation between the downmix and residual signals. It has been shown that choosing the optimal number of bits allocated between the downmix and residual signal is very essential for this scheme. A methode for efficient bit allocation is now being developed. For maintain scalability, a residual switch is accordingly applied in the decoder to select either residual or decorrelated signals. The experimental results have shown that the Improved PS scheme can outperform AAC at a bit rate of 128 kb/s. This has created a possibility of using the Improved PS as a single universal stereo audio codec for various internet applications and services.

Acknowledgment

This work was funded by the Ministry of Education and Culture, Republic of Indonesia, under DIPA Universitas Andalas, contract no. 023.04.2.415061/2014.

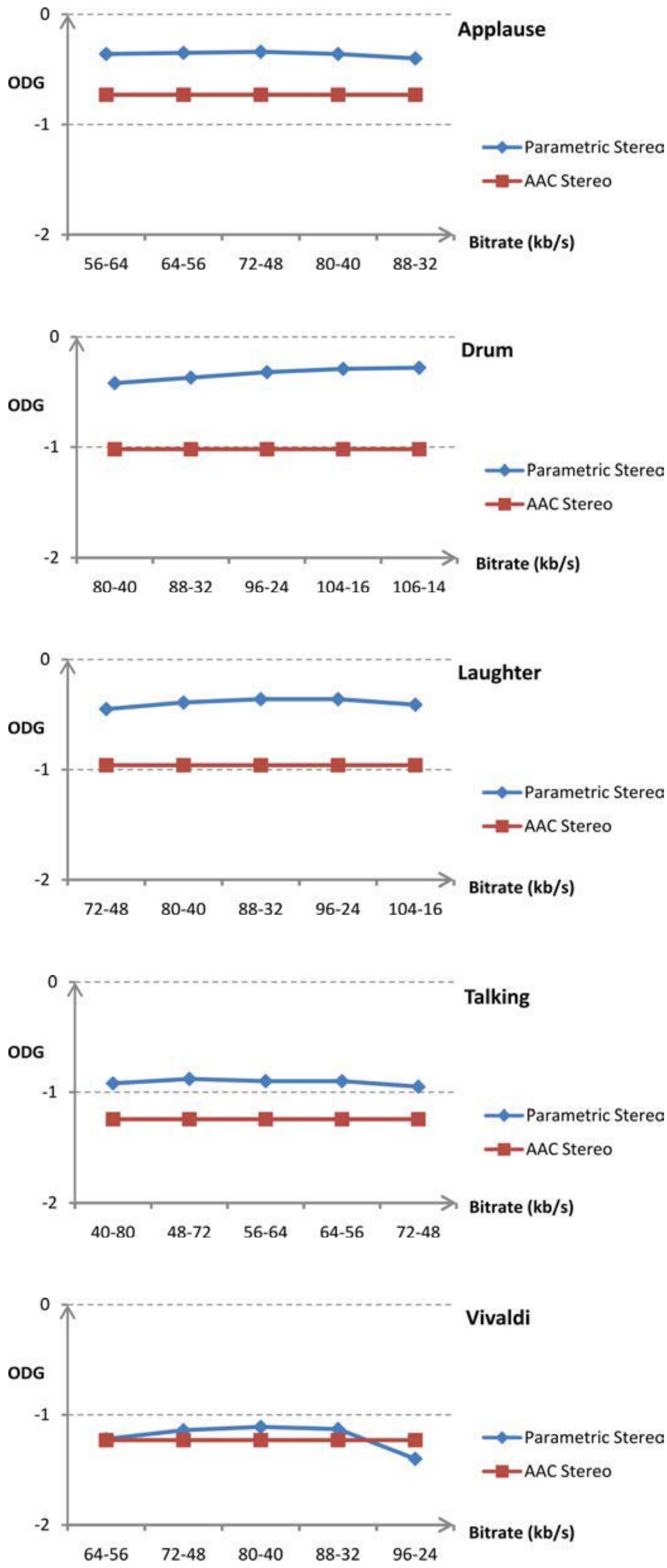

Fig. 4. ODG scores of Improved PS and AAC tested using 5 audio excerpts.

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] 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.

[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] 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.

[5] I. Elfitri, A. M. Kondoz, and B. Gunel, “Spatial audio coding,” UK Patent Publication No. GB2485979, 2012.

[6] 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.

[7] K. Brandenburg, G. Stoll, Y. Dehery, J. Johnston, L. Kerkhof, and E. Schroder, “ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio,” *J. Audio Eng. Soc.*, vol. 42 No. 10, pp. 780–792, 1994.

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

[9] ——, “Information Technology - Generic coding of moving pictures and associated audio information, Part 7: Advanced Audio Coding,” ISO/IEC 13818-7:2006(E), International Standards Organization, Geneva, Switzerland, 2006.

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

[11] 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.

[12] 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.

[13] J. Engdegard, H. Purnhagen, J. Roden, and L. Liljeryd, “Synthetic ambience in parametric stereo coding,” Presented at the 116^{th} Convention of the Audio Engineering Society, Berlin, Germany, May 2004.

[14] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, “Advances in parametric coding for high-quality audio,” Presented at the 114^{th} Convention of the Audio Engineering Society, Amsterdam, The Netherlands, Mar. 2003.

[15] 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.

[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] ISO/IEC, “Information Technology - MPEG Audio Technologies, Part 1: MPEG Surround,” ISO/IEC 23003-1:2007(E), International Standards Organization, Geneva, Switzerland, 2007.

[18] ITU-R, “Method for Objective Measurements of Perceived Audio Quality,” Recommendation ITU-R BS.1387-1, 2001.