Pilot-Based Coding Scheme for Parametric Stereo in Enhanced aacPlus

Hee-Suk Pang

ABSTRACT—*We propose a pilot-based coding (PBC) scheme for lossless bit rate reduction of parametric stereo (PS) in enhanced aacPlus. It uses PBC in addition to the existing frequency and time differential coding to encode and decode PS parameter indexes. We also design optimal Huffman codebooks (HCBs) for PBC in the proposed scheme. Experiments show that the proposed scheme is superior to the original coding scheme, where both the new coding structure and the optimal HCBs contribute to the bit rate reduction.*

Keywords—*Pilot-based coding, lossless bit rate reduction, parametric stereo, enhanced aacPlus.*

I. Introduction

Enhanced aacPlus is a state-of-the-art audio codec that guarantees good sound quality at very low bit rates [1]. It is composed of three parts: advanced audio coding (AAC), spectral band replication (SBR), and parametric stereo (PS). The PS tool is used to generate stereo signals from mono downmix signals using stereo parameters (SPs) composed mainly of inter-channel intensity differences (IIDs) and inter-channel coherence (ICC) [2]. The tool is also used for MPEG4 HE-AAC v2 [3].

In PS encoding, the SPs are extracted from input stereo signals, and their quantized indexes are differentially coded and Huffman-coded for bit stream composition. Recently, many coding methods have been proposed for bit rate reduction of MPEG Surround, which is a new standard for multi-channel audio coding [4]. Channel level differences (CLDs), which are one of the spatial parameters of MPEG Surround, were replaced by virtual source location information (VSLI) in [5] and [6]. Because VSLI essentially requires a new quantization table, it leads to lossy bit rate reduction of MPEG Surround. In [7] and [8], pilot-based coding (PBC) and extended PBC, which lead to lossless bit rate reduction of MPEG Surround, were proposed.

Based on the fact that the spatial parameters of MPEG Surround are essentially similar to the SPs of PS, we propose a PBC scheme for lossless bit rate reduction of PS in enhanced aacPlus. The proposed scheme includes a new bit stream structure and encoding/decoding processes optimized for PS. We also describe the design procedure of new Huffman codebooks (HCBs) for PBC and use them in the scheme, which is distinct from the previous work where PBC shared HCBs with the existing coding methods [7], [8].

II. Proposed Coding Scheme

In PS, the quantized indexes of IIDs and ICC have two differential coding methods before Huffman coding: frequency differential coding (FDC) and time differential coding (TDC). The discrimination between them is based on a one-bit TDC flag [3]. To minimize both the bit rate and error propagation, PS uses only FDC for refresh frames and FDC and TDC for non-refresh frames. In Fig. 1, the average bit consumption per SP set is shown for a sample stereo signal (44,100 Hz, 5 minutes), where the SP band number is 20, and refresh frames are used once per 5 frames. The results show that refresh frames require 1.18 to 1.31 times more bits than non-refresh frames. Observation of various sample signals confirms this tendency, and the rate increases up to 1.6 for some signals. The results imply that relatively fewer bits are allocated to AAC and SBR in enhanced aacPlus for refresh frames of PS, which motivates bit rate reduction for these frames.

Manuscript received May 7, 2009; revised June 3, 2009; accepted June 15, 2009.
Hee-Suk Pang (phone: +82 2 3408 3727, email: hspang@sejong.ac.kr) is with the Department of Electronics Engineering, Sejong University, Seoul, Rep. of Korea.
doi:10.4218/etrij.09.0209.0193

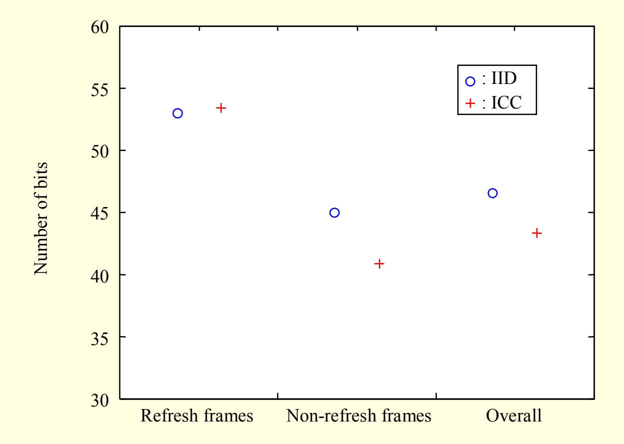

Fig. 1. Average bit consumption per SP set for a sample signal.

As a new coding method, we select PBC [7], [8]. We briefly describe it as follows. For a data set $x(n)$ whose length is N, PBC first calculates a pilot as

$$p = \arg\min \sum_{n=1}^{N} g(x(n) - p), \tag{1}$$

where p is the pilot, and $g(x)$ is the number of bits for representation of x. Then, it encodes (N+1) values, the pilot, and difference data from the pilot for bit stream composition. For PBC decoding, output values are calculated as

$$x(n) = p + x_d(n), \tag{2}$$

where $x_d(n)$ is the difference data. PBC can be used for refresh frames. It gives identical results compared to FDC and TDC because they shares quantization tables.

Based on (1) and (2), we designed preliminary HCBs for the pilots and difference data of the IIDs and ICC. The database was composed of 820,000 SP sets, which were extracted from CD samples in various genres, such as classical, pop, and world music, with a sampling frequency of 44,100 Hz. Because we used both 10 and 20 for the SP band number, the actual number of SP sets was 1,640,000.

The proposed coding scheme has three coding methods: PBC, FDC, and TDC. It utilizes only PBC and FDC for refresh frames and PBC, FDC, and TDC for non-refresh frames. Preliminary experiments showed that TDC was the most frequently used method with minimum bit consumption. As a result, we use the following bit stream structure. When a TDC flag is 1, TDC is used. When a TDC flag is 0, and a PBC flag is 1, PBC is used. Otherwise, FDC is used. Based on this structure, the flag overhead is 2 bits for FDC and PBC, and it is 1 bit for TDC. Note that the core bit streams of FDC and TDC are identical to those in the original coding scheme.

We briefly discuss the bit consumption per SP set in the proposed scheme. When FDC is used, the proposed scheme is inferior to the original scheme by 1 bit due to the flag overhead. When TDC is used, the schemes are equal. When PBC is used, the proposed scheme is superior or equal to the original scheme, where the latter case occurs due to the flag overhead.

III. Optimal HCB Design

Whereas the preliminary HCBs for PBC were created using the overall SP database sets, PBC is selectively used with FDC and TDC in the proposed scheme. Therefore, optimal HCBs should be created using SP sets where PBC is superior to FDC and TDC. We set the following two criteria for optimal HCBs:

- The bit rate of optimal HCBs should be lower than that of the preliminary HCBs for any SP band number M and any refresh frame rate L.
- Optimal HCBs should minimize the bit rate when averaged over all M and L values.

To reduce the number of candidates of optimal HCBs, we imposed a restriction that the bit lengths of the optimal HCBs should not be too different from those of the preliminary HCBs. Based on the reduced candidates, we calculated the bit rates as a function of M and L using the same database mentioned in section II. Based on the two criteria, the optimal HCBs were determined, and their bit lengths are shown in Fig. 2. The optimal HCB data can be easily derived from Fig. 2.

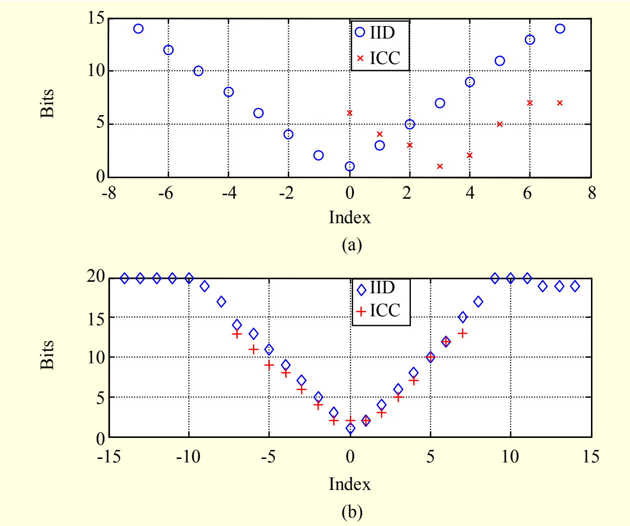

Fig. 2. Bit lengths of the optimal HCBs: (a) pilots and (b) difference data.

IV. Experimental Results

For experiments, we extracted 49 stereo items from CD samples in various genres, such as classical, pop, and soundtrack music, with a sampling frequency of 44,100 Hz. The items corresponded to 240,000 SP sets, and the CDs were

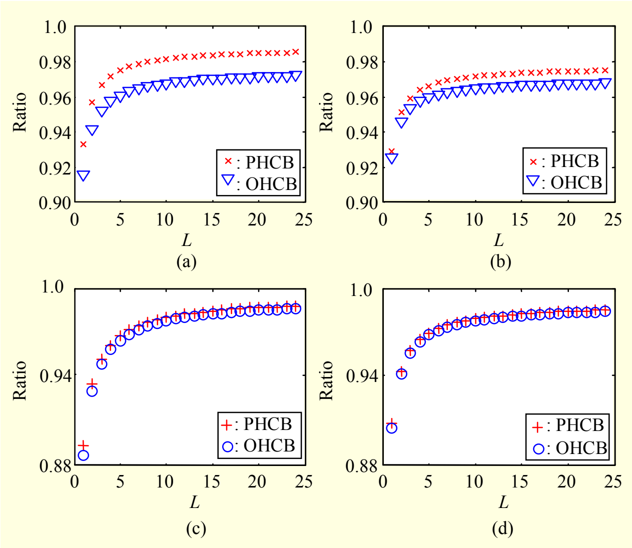

Fig. 3. Bit reduction ratios: (a) IIDs (M=10), (b) IIDs (M=20), (c) ICC (M=10), and (d) ICC (M=20).

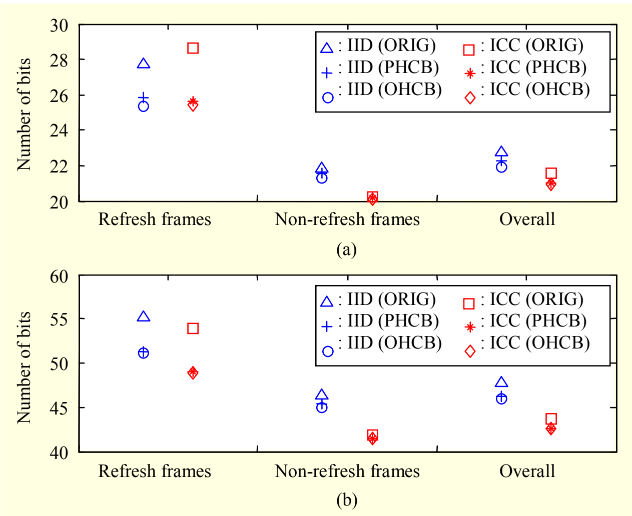

Fig. 4. Average bit consumption per SP set: (a) M=10 and (b) M=20.

different from those used in the HCB design process. Then, the signals were encoded and decoded in enhanced aacPlus format. We used two bit rates, 16 kbps and 32 kbps, where M corresponded to 10 and 20, respectively. We used L varying from 1 to 24, where L=1 implies that only refresh frames are used, and L=22 implies that refresh frames are inserted almost once per second. We show the experimental results from two viewpoints, where the results include the flag overhead as well as IID and ICC data.

In Fig. 3, the bit reduction ratios of the proposed scheme with the preliminary HCBs (PHCB) and the optimal HCBs (OHCB) to the original coding scheme (ORIG) are shown as a function of M and L. When averaged over all L values, the bit reduction ratios of OHCB to ORIG are 3.6% for M=10 and 3.8% for M=20 for IIDs. The ratios are 2.7% for M=10 and 2.6% for M=20 for ICC. The contribution of the optimal HCBs was more significant for IIDs than for ICC. The results are significant since the proposed coding scheme achieves lossless bit rate reduction of PS.

In Fig. 4, the average bit consumption per SP set is shown, where the results are averaged over all L values. The proposed scheme reduces the bit rates especially for refresh frames, where the bit reduction ratios of OHCB to ORIG are 8.5% for M=10 and 7.5% for M=20 for IIDs, and 11.3% for M=10 and 9.4% for M=20 for ICC. The results imply that the proposed scheme allocates more bits to AAC and SBR in enhanced aacPlus, especially for refresh frames of PS.

The overall PS bit stream is composed of header information as well as the flag overhead, IIDs, and ICC. Including all the information, the average bit reduction ratios of OHCB to ORIG were 8.0% for M=10 and 7.6% for M=20 for refresh frames. For non-refresh frames, the ratios were 1.47% for M=10 and 1.99% for M=20.

V. Conclusion

We proposed a new coding scheme for lossless bit rate reduction of PS in enhanced aacPlus. The proposed scheme is composed of the new coding structure and the optimal HCBs, both of which contribute to the bit rate reduction.

References

[1] 3GPP TS 26.401, *Enhanced aacPlus General Audio Codec; General Description*, v.8.0.0., 2008.

[2] 3GPP TS 26.405, *Enhanced aacPlus General Audio Codec; Encoder Specification Parametric Stereo Part*, v.8.0.0., 2008.

[3] ISO/IEC 14496-3, *Information Technology–Coding of Audio-Visual Objects–Part 3: Audio*, 2005.

[4] ISO/IEC 23003-1, *Information Technology–MPEG Audio Technologies–Part 1: MPEG Surround*, 2007.

[5] S. Beack et al., “Angle-Based Virtual Source Location Representation for Spatial Audio Coding,” *ETRI J.*, vol. 28, no. 2, Apr. 2006, pp. 219-222.

[6] K. Kim et al., “Improved Channel Level Difference Quantization for Spatial Audio Coding,” *ETRI J.*, vol. 29, no. 1, Feb. 2007, pp. 99-102.

[7] H.O. Oh, Y.W. Jung, H.S. Pang, D.S. Kim, and J. Lim, “Proposed Core Experiment on Pilot-Based Coding of Spatial Parameters of MPEG Surround,” *ISO/IEC JTC1/SC29/WG11*, M12549, 2005.

[8] H.S. Pang, J. Lim, and H.O. Oh, “Extended Pilot-Based Coding for Lossless Bit Rate Reduction of MPEG Surround,” *ETRI J.*, vol. 29, no. 1, Feb. 2007, pp. 103-106.