# **LETTER** New Context-Adaptive Arithmetic Coding Scheme for Lossless Bit Rate Reduction of Parametric Stereo in Enhanced aacPlus

Hee-Suk PANG<sup>†</sup>, Member, Jun-seok LIM<sup>†a)</sup>, and Hyun-Young JIN<sup>††</sup>, Nonmembers

**SUMMARY** We propose a new context-adaptive arithmetic coding (CAAC) scheme for lossless bit rate reduction of parametric stereo (PS) in enhanced aacPlus. Based on the probability analysis of stereo parameters indexes in PS, we propose a stereo band-dependent CAAC scheme for PS. We also propose a new coding structure of the scheme which is simple but effective. The proposed scheme has normal and memory-reduced versions, which are superior to the original and conventional schemes and guarantees significant bit rate reduction of PS. The proposed scheme can be an alternative to the original PS coding scheme at low bit rate, where coding efficiency is very important.

*key words:* context-adaptive arithmetic coding, lossless bit rate reduction, parametric stereo, coding efficiency

## 1. Introduction

Parametric stereo (PS) is a coding technique that reconstructs a stereo signal using its mono downmix signal and stereo parameters (SPs) [1], [2]. Since PS provides good sound quality with a small amount of information for SPs, it has widely been used for low bit rate audio coding, e.g. enhanced aacPlus [3] and high-efficiency advanced audio coding version 2 (HE-AAC v2) [4]–[6]. Its expanded versions are also used in MPEG Surround [7], which is multichannel audio coding, and unified speech and audio coding (USAC) [8].

PS is usually used with a core codec that encodes and decodes the downmix signal. For example, advanced audio coding (AAC) is used as the core codec in [3]–[6], [8]. The bit rate reduction of PS implies that the core codec can use more bits, which is desirable especially at low bit rate. Several methods have been proposed for bit rate reduction of PS and MPEG Surround. The channel level differences in MPEG Surround were replaced by virtual source location information with a bit rate reduction of 4.1% to 6.5% [9]. However, their proposal was not for lossless bit rate reduction. A pilot-based coding (PBC) scheme was proposed for lossless bit rate reduction of only around 2% [10]. Recently, context-adaptive arithmetic coding (CAAC) was adopted for USAC for efficient entropy coding of spectral data [8]. Then, CAAC

schemes were proposed for lossless bit rate reduction of MPEG Surround [11] and PS [12]. The CAAC schemes showed a considerable amount of bit rate reduction, e.g. 6.41% to 9.67% for the overall PS bit streams [12].

In this letter, we propose a new CAAC scheme for further bit rate reduction of PS compared with the original and conventional PS coding schemes. The proposed scheme is also for lossless bit rate reduction and guarantees identical results to those of the original PS coding scheme.

## 2. Brief Review of the Original PS Coding Scheme

This section describes a brief review of the original PS coding scheme in enhanced aacPlus [3]. PS in enhanced aacPlus uses two kinds of SPs, inter-channel intensity differences (IIDs) and inter-channel coherence (ICC). On the encoder side, an input signal is transformed into the hybrid subband domain and the SPs are extracted on a stereo band (SB) basis. The SBs are composed of one or more hybrid subbands, which are mapped to frequency. The number of SBs, *M*, is set to either 10 or 20 depending on bit rate. The SPs are quantized so that the quantized indexes range from -7 to 7 for IIDs and from 0 to 7 for ICC.

The indexes are coded by either frequency differential coding (FDC) or time differential coding (TDC) and then Huffman-coded for bit stream composition. For refresh frames where no information is utilized from previous frames, only FDC is used. For non-refresh frames, FDC and TDC are compared and a superior method is used.

## 3. The Proposed Method

#### 3.1 Analysis of Probabilities

We first built a database composed of stereo samples ripped from CDs in various genres, including classical, pop, and world music. With a sampling frequency of 44,100 Hz, the database corresponded to 822,171 frames, corresponding to 10 hours and 36 minutes. We encoded the samples using the reference enhanced aacPlus encoder [13] at the bit rates of 20 kbps and 32 kbps, where *M* corresponded to 10 and 20 at each bit rate, and extracted quantized IID and ICC indexes. Since the indexes are calculated for each SB, the database corresponds to 8,221,710 SP sets for M = 10 and 16,443,420 SP sets for M = 20. We refer to this database as the training database.

Manuscript received June 21, 2018.

Manuscript revised August 27, 2018.

Manuscript publicized September 18, 2018.

<sup>&</sup>lt;sup>†</sup>The authors are with the Department of Electrical Engineering, Sejong University, Seoul 05006, Korea.

<sup>&</sup>lt;sup>††</sup>The author is with the Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Korea.

a) E-mail: jslim@sejong.ac.kr (Corresponding author)

DOI: 10.1587/transinf.2018EDL8135



**Fig. 1** Probabilities of the SP indexes for each SB (M = 20). The circles represent the probabilities averaged over all the SBs. (a) IIDs. (b) ICC.

Using the training database, we analyzed the probabilities of the SP indexes and depicted them for each SB for M = 20 in Fig. 1. In Fig. 1 (a), the results of each SB are similar in shape, but their values at index = 0 range from 0.27 to 0.50 depending on SB. In Fig. 1 (b), the results also vary according to SB. Results showed similar tendencies for M = 10. The original and conventional PS coding schemes are based on the SP indexes' probabilities averaged over all the SBs [3], [10], [12]. However, we utilize the probabilities of each SB instead of the averaged probabilities.

#### 3.2 Determination of the PS Coding Scheme

In context-adaptive coding (CAC), a source is coded according to a context. We apply CAC to PS so that an SP index is coded according to its adjacent SP indexes. Suppose that a source x(k, n) is an SP index, where k is an SB index for  $1 \le k \le M$  and n is a time index. The original PS coding scheme is based on the assumption that x(k, n) does not change abruptly both in the SB domain and in the time domain [3]. We also utilize the assumption and use x(k - 1, n)and x(k, n - 1) as contexts, which are adjacent to x(k, n) in the SB and time domains. Note that x(k+1, n) and x(k, n+1)are also adjacent to x(k, n), but they are unavailable when we decode x(k, n).

We consider three CAC methods as shown in Fig. 2. Suppose that x(k, n) is a source. The first method is using x(k - 1, n) as a context for  $2 \le k \le M$ , which we refer to as SB CAC (SB-CAC). Exceptionally, x(1, n) should be coded without contexts. The second method is using x(k, n - 1) as a context for  $1 \le k \le M$ , which we refer to as time CAC (T-CAC). The third method is using both x(k - 1, n) and



Fig. 2 Three CAC methods. (a) SB-CAC. (b) T-CAC. (c) SBT-CAC.

 Table 1
 Percentages of frames where each CAAC method uses the least bits for either IIDs or ICC.

Mathad	IIDs	ICC	IIDs	ICC
Method	(M = 10)	(M = 10)	(M=20)	(M = 20)
SB-CAAC	17.14%	13.29%	13.38%	11.31%
T-CAAC	44.83%	40.85%	31.11%	28.36%
SBT-CAAC	66.93%	70.18%	70.67%	73.20%

x(k, n - 1) as contexts for  $2 \le k \le M$ , which we refer to as SB-time CAC (SBT-CAC). Exceptionally, x(1, n) should be coded only with the context x(1, n - 1).

We implemented SB-dependent CAAC for three methods using the training database, where CAC was combined with arithmetic coding whose cumulative probability tables were generated for each SB. Arithmetic coding codes a sequence of symbols into a single number in the unit interval [0, 1) and is more efficient than Huffman coding [14]. For an integer implementation of arithmetic coding, we used 16-bit unsigned format for the tables. In case the number of occurrences was 0, which sometimes occurred when the source was unusually different from the context, it was replaced by 1 before generating the tables. For detailed descriptions of the integer implementation of arithmetic coding, refer to [14].

To compare three CAAC methods, we performed two preliminary experiments using the training database. The first experiment was directly comparing them for either IIDs or ICC. We calculated their bit consumption on a frame basis and calculated the percentages of frames where each method uses the least bits. In Table 1, SBT-CAAC uses the least bits most often. The second experiment was comparing their combinations for both IIDs and ICC. We calculated the percentages of frames where each combinative method for IIDs and ICC uses the least bits. Let us denote a combinative method that uses 'A'-CAAC for IIDs and 'B'-CAAC for ICC as 'A-B' CAAC. In Table 2, SBT-SBT CAAC, T-SBT CAAC, and SBT-T CAAC use the least bits often. In the calculation of Tables 1 and 2, we counted the frame numbers of multiple methods if they used the same least bits.

Considering the results in Table 2, we design a PS cod-

 Table 2
 Percentages of frames where each combinative method uses the least bits for both IIDs and ICC.

Method	M = 10	M = 20
SB-SB CAAC	4.87%	4.20%
SB-SBT CAAC	10.49%	8.39%
SB-T CAAC	5.04%	2.13%
SBT-SB CAAC	7.39%	6.53%
SBT-SBT CAAC	47.80%	53.10%
SBT-T CAAC	26.54%	19.58%
T-SB CAAC	4.20%	2.07%
T-SBT CAAC	30.67%	22.13%
T-T CAAC	19.71%	10.81%

 Table 3
 Average bit consumption (bits/frame) of three CAAC schemes.

CAAC scheme	M = 10	M = 20
Scheme 1	37.17	76.36
Scheme 2	37.60	76.64
Scheme 3	37.87	77.03



Fig. 3 The proposed PS coding scheme.

ing scheme. A SP set is composed of a flag and SP indexes in the bit stream, where the indexes are coded by the most efficient method among predefined CAAC methods and the flag specifies which method was used. If the length of the flag is m bits,  $2^m$  CAAC methods are available. We consider both functionality and efficiency to determine the scheme. From the viewpoint of functionality, the scheme should support refresh frames to prevent error propagation and to provide random access functionality, which is important for communication and streaming services. From 9 combinative methods in Table 2, SB-SB CAAC is the only method that can be used for refresh frames and must be included in the scheme irrespective of its efficiency. For efficiency, we performed additional experiments comparing three schemes. Scheme 1 uses a 1-bit flag and is composed of SB-SB and SBT-SBT CAAC methods. Scheme 2 uses a 2-bit flag and is composed of SB-SB, SBT-SBT, T-SBT, and SBT-T CAAC methods. Scheme 3 uses a 2-bit flag and is composed of SB-SB, SBT-SBT, SB-SBT, and SBT-SB CAAC methods, whose structure was used in [12]. Note that all the schemes include SB-SB CAAC to support refresh frames. We calculated the bit consumption of each scheme for SP sets using the training database. In Table 3, scheme 1 uses the least bits. Since scheme 1 is simple in its structure but effective in coding efficiency, it will be used for the proposed scheme.

## 3.3 The Proposed PS Coding Scheme

The proposed PS coding scheme is shown in Fig. 3, which is composed of two coding methods. The first method is to apply SB-SB CAAC to SP indexes, i.e. SB-CAAC to both IIDs

Table 4Table sizes of the proposed scheme in word (1 word=16 bits).

		-	
Coding method	IID	ICC	
FAC $(k = 1)$	15	8	
SB-CAAC $(2 \le k \le M)$	$(M - 1) \times 15^2$	$(M - 1) \times 8^2$	
T-CAAC $(k = 1)$	15 <sup>2</sup>	$8^{2}$	
SBT-CAAC $(2 \le k \le M)$	$(M - 1) \times 15^3$	$(M-1) \times 8^3$	
Overall	$M \times 4176 - 3864$		

and ICC. Note that fixed arithmetic coding (FAC) is used for the first SB. The second method is to apply SBT-SBT CAAC to SP indexes, i.e. SBT-CAAC to both IIDs and ICC. Note that T-CAAC is used for the first SB. On the encoder side, both methods are compared and a superior method is used for non-refresh frames. For refresh frames, the first method is used. On the decoder side, the flag is first read and the decoding method is determined.

The proposed scheme has normal and memory-reduced versions. The normal version has the advantage over the previous scheme [12] from two viewpoints. Firstly, it is an SB-dependent CAAC scheme whose cumulative probability tables are defined for each SB as explained in Sect. 3.1, which is not the case in [12]. It achieves the bit rate reduction at the cost of memory size as shown in Table 4, which is acceptable in most cases due to recent development of storage devices. The table sizes are 37,896 words for M = 10 and 79,656 words for M = 20 in Table 4, where 1 word means 16 bits. Secondly, the coding structure of the proposed scheme is newly determined based on the experiments used for Tables 1, 2, and 3. It is simpler but more effective than that of the previous scheme [12].

In specific applications where the memory size is limited, the memory-reduced version is useful. It also uses the new coding structure, which is different from that of the previous scheme. However, it does not utilize the SBdependent CAAC scheme; its cumulative probability tables for CAAC are shared by all SBs. Then, its table size is 4,488 words, which is identical to that of the previous scheme [12].

#### 4. Experimental Results

We built a database composed of stereo samples ripped from CDs in various genres, including classical, pop, and sound-track music with a sampling frequency of 44,100 Hz. The samples corresponded to 238,956 frames, corresponding to 3 hours and 5 minutes, and were different from those used for the training database. We encoded the samples using the reference enhanced aacPlus encoder [13] at the bit rates of 20 kbps and 32 kbps, where *M* corresponded to 10 and 20 respectively, and extracted the IID and ICC indexes. We refer to this database as the test database.

We applied the original PS coding scheme [3], PBC scheme [10], CAAC scheme [12], and proposed scheme to the test database and calculated the average bit consumption per frame. The refresh rate R is defined so that refresh frames are inserted every R second and was set to 1. In Table 5, the results are listed for SP sets, which are composed of the flag and SP indexes, and for the overall PS bit

	Scheme	SP sets	SP sets	SP sets	Overall
М		(refresh	(non-refresh	(overall	PS bit
		frames)	frames)	frames)	streams
	Original	55.59	41.62	42.29	46.71
10	PBC	50.26	40.98	41.42	45.85
		(9.59%)	(1.54%)	(2.06%)	(1.84%)
	CAAC	48.19	38.63	39.08	43.51
		(13.31%)	(7.18%)	(7.59%)	(6.85%)
	Proposed	47.19	37.91	38.35	42.78
	(reduced)	(15.11%)	(8.91%)	(9.32%)	(8.41%)
	Proposed	46.19	37.20	37.62	42.05
	(normal)	(16.91%)	(10.62%)	(11.04%)	(9.98%)
	Original	108.38	87.54	88.53	92.96
	PBC	99.30	85.74	86.38	90.81
		(8.38%)	(2.06%)	(2.43%)	(2.31%)
	CAAC	94.55	78.47	79.24	83.67
20		(12.76%)	(10.36%)	(10.49%)	(9.99%)
	Proposed	93.55	77.91	78.66	83.09
	(reduced)	(13.68%)	(11.00%)	(11.15%)	(10.62%)
	Proposed	90.80	76.23	76.92	81.35
	(normal)	(16.22%)	(12.92%)	(13.11%)	(12.49%)

 Table 5
 Average bit consumption (bits/frame) of each scheme for the refresh rate is 1. The numbers in parentheses represent the bit reduction ratios of each scheme to the original PS coding scheme.

streams, which are composed of all the information in the PS bit streams including the SP sets. The normal version of the proposed scheme is superior to the original and conventional schemes and achieves significant bit reductions. Compared to the original scheme, the bit rate reductions of the normal version correspond to 100 bps for M = 10 and 250 bps for M = 20. The memory-reduced version also shows meaningful improvements over the original and conventional schemes. Its bit rate reductions correspond to 85 bps for M = 10 and 213 bps for M = 20.

Since the proposed scheme guarantees lossless bit rate reduction of PS, its bit streams contain identical information to that of the original PS coding scheme. Enhanced aacplus is composed of AAC, spectral band replication, and PS [15], and the saved bits in PS can be used for the core codec, AAC. For supplementary information, we calculated the bit reduction ratios of the proposed scheme to the original scheme for the entire enhanced aacPlus bit streams. In the experiments, the PS bit streams at the bit rates of 20 kbps and 32 kbps. Then, the bit reduction ratios in Table 5 corresponded to 0.50% at 20 kbps and 0.78% at 32 kbps for the entire enhanced aacPlus bit streams.

In addition, the bit reduction ratios of each scheme to the original scheme are depicted for the overall PS bit streams as a function of refresh rate R in Fig. 4. The results show that the proposed scheme guarantees significant and consistent lossless bit reduction of PS in all cases.

#### 5. Conclusion

We proposed a new CAAC scheme for lossless bit rate reduction of PS in enhanced aacPlus. The experimental results show that it achieves significant and consistent lossless bit rate reduction of PS. The saved bits by our scheme can be



**Fig.4** Bit reduction ratios of each scheme to the original PS coding scheme as a function of refresh rate. The results were calculated for the overall PS bit streams. (a) M = 10. (b) M = 20.

used for the core codec, which is desirable at low bit rate.

Since PS provides good sound quality with a small amount of information, it is commonly used for low bit rate audio coding technologies [3]–[8]. Our scheme, which is not compatible with the existing international standards, can be one possible alternative to the original PS coding scheme when we develop a new audio coding technology operating at low bit rate, where coding efficiency is very important.

#### References

- J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," EURASIP J. Appl. Signal Process., vol.2005, no.9, pp.1305–1322, June 2005.
- [2] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, "Low complexity parametric stereo coding," AES Convention, Paper Number 6073, Berlin, Germany, May 2004.
- [3] 3GPP TS 26.405, Enhanced aacPlus general audio codec; encoder specification parametric stereo part, v.14.0.0., March 2017.
- [4] ISO/IEC 14496-3, Information technology coding of audio-visual objects - part 3: Audio, subpart 8: technical description of parametric coding for high quality audio. 2009.
- [5] S. Meltzer and G. Moser, "MPEG-4 HE-AAC v2 audio coding for today's digital media world," EBU Technical Review, Jan. 2006.
- [6] J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding," IEEE Signal Process. Mag., vol.25, no.3, pp.137–142, May 2008.
- [7] ISO/IEC 23003-1, Information technology MPEG audio technologies - part 1: MPEG Surround, 2007.
- [8] ISO/IEC 23003-3, Information technology MPEG audio technologies - part 3: Unified Speech and Audio Coding, 2012.
- [9] S. Beack, J. Seo, H. Moon, K. Kang, and M. Hahn, "Angle-based virtual source location representation for spatial audio coding," ETRI J., vol.28, no.2, pp.219–222, April 2006.
- [10] H.-S Pang, "Pilot-based coding scheme for parametric stereo in enhanced aacPlus," ETRI J., vol.31, no.5, pp.613–615, Oct. 2009.
- [11] S. Yoon, H.-S. Pang, and K.-M. Sung, "Context-adaptive arithmetic coding scheme for lossless bit rate reduction of MPEG Surround in USAC," IEICE Trans. Inf. & Syst., vol.E95-D, no.7, pp.2013–2016, July 2012.
- [12] H.S. Pang, J.S. Lim, and S. Lee, "Context-based arithmetic coding scheme for parametric stereo in enhanced aacPlus," J. Inf. Sci. Eng., vol.30, no.6, pp.2019–2032, Nov. 2014.
- [13] 3GPP TS 26.410, Enhanced aacPlus general audio codec; floating-

point ANSI-C code, v.14.0.0., March 2017.

- [14] K. Sayood, Introduction to Data Compression, Morgan Kaufmann Publishers, San Francisco, pp.81–115, 2005.
- [15] 3GPP TS 26.401, Enhanced aacPlus general audio codec; general description, v.14.0.0., March 2017.