LETTER

Context-Adaptive Arithmetic Coding Scheme for Lossless Bit Rate Reduction of MPEG Surround in USAC

Sungyong YOON[†,†††a)], ***Nonmember***, **Hee-Suk PANG**[††b)], ***and*** **Koeng-Mo SUNG**[†††], ***Members***

SUMMARY We propose a new coding scheme for lossless bit rate reduction of the MPEG Surround module in unified speech and audio coding (USAC). The proposed scheme is based on context-adaptive arithmetic coding for efficient bit stream composition of spatial parameters. Experiments show that it achieves the significant lossless bit reduction of 9.93% to 12.14% for spatial parameters and 8.64% to 8.96% for the overall MPEG Surround bit streams compared to the original scheme. The proposed scheme, which is not currently included in USAC, can be used for the improved coding efficiency of MPEG Surround in USAC, where the saved bits can be utilized by the other modules in USAC.

key words: *lossless bit rate reduction, MPEG Surround, USAC, context-adaptive arithmetic coding*

1. Introduction

In MPEG audio, channel expansion techniques are used to improve the sound quality at low bit rates. For example, parametric stereo (PS) reconstructs stereo signals using mono downmix signals and stereo parameters [1]. MPEG Surround reconstructs multi-channel signals using mono or stereo downmix signals and spatial parameters [2]. A modified version of MPEG Surround is also adopted as a channel expansion tool in unified speech and audio coding (USAC), which is in the standardization process in MPEG, for the consistent coding quality of both speech and audio signals at very low bit rate [3].

Since MPEG Surround is used with a core codec which encodes and decodes downmix signals, the bit reduction in MPEG Surround implies that the core codec can use more bits, which can improve the sound quality of reconstructed signals. Recently, methods have been proposed for bit rate reduction of MPEG Surround, where the targets are mainly 5.1 channel signals. For example, the channel level differences among spatial parameters are replaced by virtual source location information, which results in a lossy bit rate reduction of around 5% [4]. Extended pilot-based coding results in a lossless bit rate reduction of less than 3% for each spatial parameter [5]. Since both methods are based on the original coding scheme of MPEG Surround, which uses context-independent Huffman coding, their bit rate reductions are relatively small.

Recently, a context-adaptive arithmetic coding scheme exploiting the time-frequency dependencies among MDCT coefficients has been proposed for the efficient coding of spectral data in USAC [3], [6]. In this letter, we propose a new coding scheme for lossless bit rate reduction of MPEG Surround in USAC, which replaces the existing Huffman coding with context-adaptive arithmetic coding for efficient coding of spatial parameters. Our target is to achieve a significant bit reduction of the MPEG Surround module, where the saved bits can be used by the other modules in USAC.

2. Determination of Contexts

In MPEG Surround encoding, a time-domain signal is transformed into the hybrid subband domain and then spatial parameters are calculated [2]. In MPEG Surround in USAC, the spatial parameters are composed of channel level differences (CLDs), inter-channel correlation (ICC), and inter-channel phase differences (IPDs), which represent the level differences, correlation, and phase differences between channels on a parameter band basis, where the parameter band is composed of one or more hybrid subbands [2], [3]. Whereas CLDs and ICC are always used, IPDs are used mainly at high bit rate. The spatial parameters are quantized so that the quantized indexes are integer values ranging from -15 to 15 for CLD, from 0 to 7 for ICC, and from 0 to 15 for IPD [2], [3].

The quantized indexes are further coded by lossless coding tools, Huffman coding and grouped PCM coding [3]. In the former, the differential index values are calculated either in the time or parameter band domain and then Huffman-coded either on an index basis or on a pair of indexes basis. The latter, grouped PCM coding, mainly provides functionality which limits the maximum bit consumption in an extraordinary case. Since one of the main target applications of USAC is broadcasting, it is important to support random access functionality and to prevent the error propagation. This can be done by periodically inserting refresh frames in which no information from the previous frames is utilized. In MPEG Surround in USAC, only frequency differential coding and grouped PCM coding are used for refresh frames, where the term 'frequency' is used since a parameter band is mapped to frequency. For non-refresh frames, all the coding tools including time differential coding are applicable.

Manuscript received December 16, 2011.
Manuscript revised March 8, 2012.

[†]The author is with Convergence Research Lab., LG Electronics, Seoul 137–130, Korea.
[††]The author is with the Dept. of Electronics Eng., Sejong University, Seoul 143–747, Korea.
[†††]The authors are with School of Electrical Eng. and Computer Science, Seoul National University, Seoul 151–744, Korea.
a) E-mail: rt60@paran.com
b) E-mail: hspang@sejong.ac.kr (Corresponding author)

DOI: 10.1587/transinf.E95.D.2013

In context-adaptive coding, a source is coded according to a context, where the source and the context are the quantized indexes of spatial parameters in MPEG Surround in USAC. To replace the context-independent coding scheme of MPEG Surround with a context-adaptive coding scheme, the optimal context should be first determined. In Fig. 1, the quantized indexes of spatial parameters are shown in the time-parameter band domain, where $X(n,k)$ is a quantized index with a time index n and a parameter band index k. Since the indexes are decoded in order of parameter band and time, the contexts available for a source $X(n,k)$ are limited to the white region in Fig. 1. In Fig. 2, the probabilities of the differential quantized indexes of CLDs and ICC are depicted for a number of stereo samples with the number of parameter bands M=20. The results show that the probabilities that the differential indexes are around 0 are high, which means that the indexes adjacent to each other in the time-parameter band domain are very similar. Further experiments show that this tendency is consistent for other values of M. Therefore, we use $X(n-1,k)$ and $X(n,k-1)$ as contexts for a source $X(n,k)$, which are the most adjacent to $X(n,k)$ in the time-parameter band domain. Based on a combination of $X(n-1,k)$ and $X(n,k-1)$, there are three context-adaptive coding methods as follows.

First, time-frequency context (TF-context) coding utilizes both $X(n-1,k)$ and $X(n,k-1)$. A problem when simply using a pair of $X(n-1,k)$ and $X(n,k-1)$ as a context and $X(n,k)$ as a source is that it requires a large table size for probability tables. For example, TF-context coding for CLD requires 31^3=29,791 values for probability tables since it requires $X(n-1,k)$, $X(n,k-1)$, and $X(n,k)$ whose CLDs all have 31 quantized index values.

An alternative is to use differential indexes for both the context and the source. Then the context is either $X(n-1,k)-X(n,k-1)$ or $X(n,k-1)-X(n-1,k)$. We can select either of them since they are practically identical from the viewpoint of probability. Since $X(n-1,k)$ and $X(n,k-1)$ are already available, the source is either $X(n,k)-X(n-1,k)$ or $X(n,k)-X(n,k-1)$ to decode $X(n,k)$. If we assume that $X(n-1,k)=a$ and $X(n,k-1)=b$, the probability distribution of $X(n,k)-a$ is only the shifted version of the probability distribution of $X(n,k)-b$. Therefore, both cases lead to the equal coding performance and we can select either of them. In this letter, we use $X(n-1,k)-X(n,k-1)$ as the context and $X(n,k)-X(n,k-1)$ as the source. This modified version of TF-context coding utilizes all the three indexes, $X(n-1,k)$, $X(n,k-1)$, and $X(n,k)$, reducing the table size significantly. For example, the table size for CLD decreases from 29,791 values to 61^2=3,721 values.

Second, time context (T-context) coding utilizes $X(n-1,k)$ as the context and $X(n,k)$ as the source. Therefore, it requires no information from previous parameter bands.

Finally, frequency context (F-context) coding utilizes $X(n,k-1)$ as the context and $X(n,k)$ as the source. Therefore, it requires no information from previous frames.

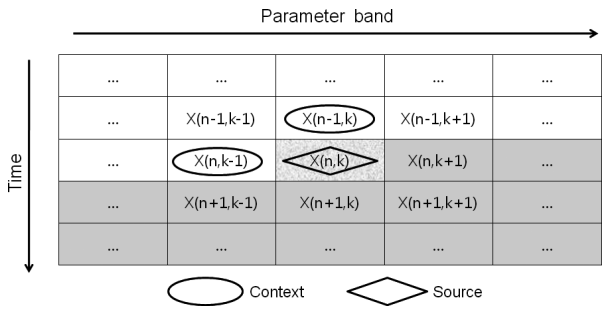

Fig. 1 The quantized indexes of spatial parameters in the time-parameter band domain.

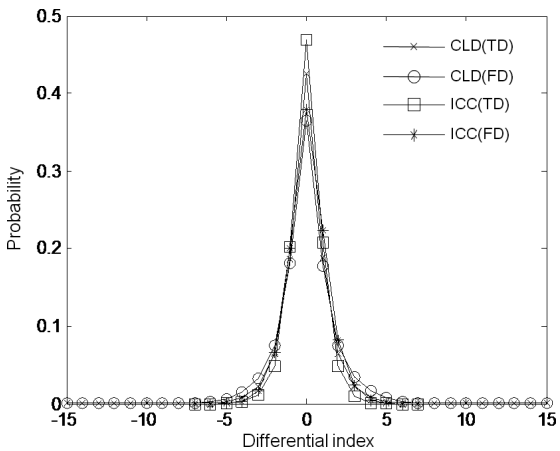

Fig. 2 The probabilities of the time differential (TD) and frequency differential (FD) quantized indexes of CLDs and ICC for M=20.

3. The Proposed Coding Scheme

In Fig. 3, the quantized indexes are divided into 4 regions in the time-parameter band domain, where the gray region corresponds to a refresh frame. The applicability of the context-adaptive coding methods depends on the region to which the quantized index of a spatial parameter belongs. Region B corresponds to parameter bands for $2 \le k \le M$ for non-refresh frames, where M is the number of parameter bands, and composes the majority of the overall quantized indexes. All the three coding methods, F-context, T-context, and TF-

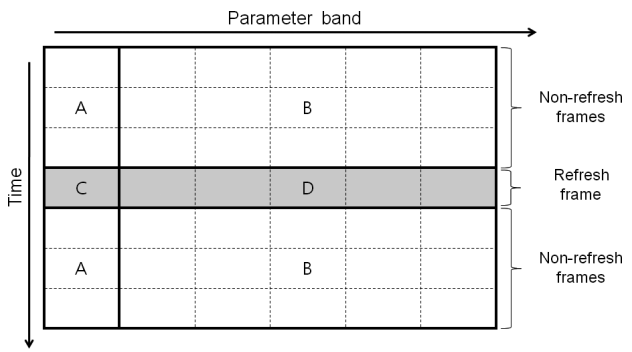

Fig. 3 4 divided quantized index regions in the time-parameter band domain.

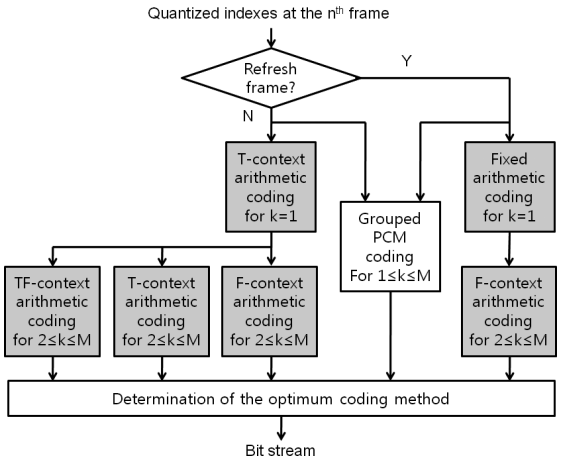

Fig. 4 Encoding scheme of the proposed method, where the gray areas represent the proposed parts.

context coding, are applicable for this region. Region A corresponds to the first parameter band for non-refresh frames, where only T-context coding is applicable. Region D corresponds to parameter bands for $2 \le k \le M$ for refresh frames, where only F-context coding is applicable. Region C corresponds to the first parameter band for refresh frames, where no context coding is applicable. Therefore we should apply a context-independent coding for this region, which we refer to as fixed coding.

The proposed method is based on arithmetic coding in conjunction with context adaptation. Arithmetic coding codes a sequence of symbols into a single number in the unit interval [0, 1) [7] and is more appropriate for context adaptation than Huffman coding. The cumulative probability tables for each context were generated using 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, where the total number of frames was about 746,000. For an integer implementation of arithmetic coding, the number of bits for the [0, 1) interval should be larger than the number of bits for cumulative probability tables by 2 bits [7]. Therefore, we used 16-bit and 14-bit format for the interval and the probability tables. The tables for the fixed arithmetic coding were also generated using the database.

In Fig. 4, the encoding scheme of the proposed method is shown, which is applied on a spatial parameter basis. The scheme is based on the applicability of the context-adaptive coding methods. Additionally, grouped PCM coding is used for all cases to prevent extreme bit consumption as in the original MPEG Surround. The number of coding methods is 2 and 4 for refresh and non-refresh frames, where 1-bit and 2-bit flags are required for each case. For decoding, the flag is first decoded and the indexes are decoded according to the coding method. The proposed method guarantees identical results to those of the original MPEG Surround in USAC since only the context-independent Huffman coding tool is replaced by the context-adaptive arithmetic coding tool.

Table 1 Table size of the proposed method in word (1 word=16 bits).

Coding method	CLD	ICC	IPD
F-context	31∗31	8∗8	16∗16
T-context	31∗31	8∗8	16∗16
TF-context	31∗31	15∗15	16∗16
Fixed	31	8	16
Overall	4,059		

The table size of the proposed method is listed in Table 1, where each value represents the number of context values multiplied by the number of source values. Exceptionally TF-context coding for CLD is applied only to the central values (−15 to 15) of differential CLD indexes for both the context and the source, which originally range from −30 to 30, to reduce the table size. This reduced version covers more than 99.9% of the differential CLD indexes in the database, where the outliers are coded by the other coding methods. In case of IPD, the number of differential index values is 16 due to the modulo properties of the phase.

4. Experimental Results

For evaluation, we compare the bit consumption of the proposed method with that of the original MPEG Surround in USAC, where the latter corresponds to the recent reference software, the reference model 11 (RM11) [8]. The database for experiments is composed of samples ripped from CDs in various genres, including classical, pop, and soundtrack music with a sampling frequency of 44,100 Hz, which is different from the database used for generating the cumulative probability tables. The total number of frames is about 217,000.

We encoded the signals in the database using RM11 at bit rates of 16, 20, 24, and 32 kbps and calculated bit consumption for each spatial parameter and the overall MPEG Surround bit stream. Then the MPEG Surround bit stream was recomposed according to the proposed method and its bit consumption was calculated for each parameter and the overall bit stream. The numbers of parameter bands were 10 and 10 for CLD and ICC at bit rates of 16 kbps and 20 kbps, where the IPD was not used at the low bit rates, and 20, 20, and 10 for CLD, ICC, and IPD at bit rates of 24 kbps and 32 kbps. In Table 2, the bit reduction ratios of the proposed method to RM11 are listed for the entire database, where refresh frames are inserted every one second. The bit reduction ratios range from 9.93% to 12.14% for spatial parameters. The bit reduction ratios range from 8.64% to 8.96% for the overall MPEG Surround bit stream, which contains header and common information as well as the spatial parameters. The results show that the proposed method achieves a significant lossless bit rate reduction of MPEG Surround. In the experiments, the MPEG Surround bit streams comprised 5.39%, 4.76%, 9.89%, and 8.66% of the entire USAC bit streams at 16, 20, 24, and 32 kbps. As a result, the bit reduction ratios in Table 2 correspond to 0.48%, 0.42%, 0.86%, and 0.75% at each bit rate for the entire USAC bit streams.

Table 2 Average bit consumption and bit reduction ratios of the proposed method to RM11 with refresh frames inserted every one second.

Bit rate (kbps)	Parameter	RM11 (bits/frame)	The proposed method (bits/frame)	Bit reduction ratio (%)
16	CLD	24.85	22.11	11.05
	ICC	22.82	20.13	11.80
	Overall	60.67	55.23	8.96
20	CLD	24.71	22.00	10.98
	ICC	22.60	19.99	11.54
	Overall	60.31	54.99	8.82
24	CLD	49.21	44.28	10.01
	ICC	43.56	38.27	12.14
	IPD	20.72	18.66	9.95
	Overall	140.49	128.22	8.74
32	CLD	48.65	43.80	9.97
	ICC	42.88	37.76	11.95
	IPD	20.65	18.60	9.93
	Overall	139.18	127.16	8.64

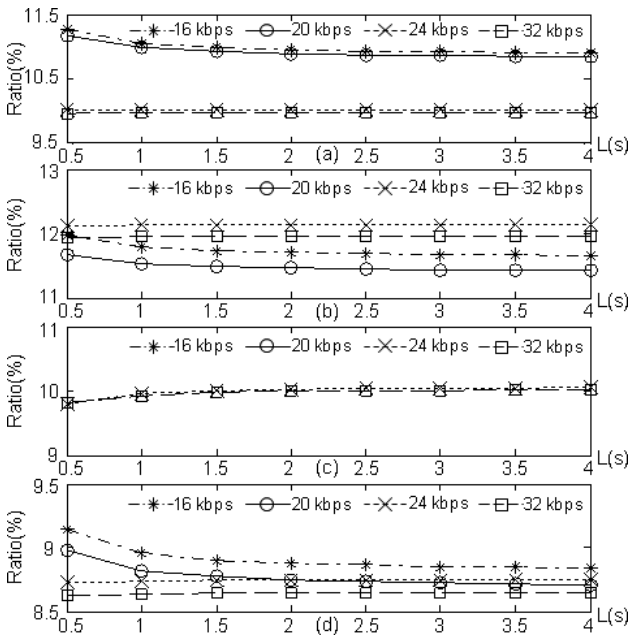

Fig. 5 Bit reduction ratios of the proposed method to RM11 as a function of a refresh frame rate L. (a) CLD, (b) ICC, (c) IPD, and (d) Overall.

Additionally, the bit rate reduction ratios of the proposed method to RM11 are depicted as a function of a refresh frame rate L in Fig. 5, where a refresh frame is inserted every L second. The results are depicted for each spatial parameter and the overall MPEG Surround bit stream in USAC, which shows that the proposed method guarantees a significant and consistent lossless bit rate reduction irrespective of L.

5. Conclusion

We propose a new coding scheme based on context-adaptive arithmetic coding for lossless bit rate reduction of MPEG Surround in USAC. Simulation results show that the proposed method achieves a significant lossless bit rate reduction. The proposed method, which is not currently included in USAC, can be used to improve the coding efficiency of MPEG Surround in USAC.

References

[1] 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.
[2] ISO/IEC 23003-1, Information technology—MPEG audio technologies—part 1: MPEG Surround, 2007.
[3] ISO/IEC JTC1/SC29/WG11, FDIS of unified speech and audio coding, N12231, Sept. 2011.
[4] 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.
[5] 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, pp.103–106, Feb. 2007.
[6] G. Fuchs, M. Multrus, M. Neuendorf, and R. Geiger, "MDCT-based coder for highly adaptive speech and audio coding," Proc. EUSIPCO 2009, pp.1264–1268, Aug. 2009.
[7] K. Sayood, Introduction to data compression, Morgan Kaufmann Publishers, San Francisco, pp.81–115, 2005.
[8] http://wg11.sc29.org/svn/repos/MPEG-D/trunk/USAC/USAC_RM11_ReferenceSoftware_20110704.zip