Audio Transcoding for Audio Streams from a T-DTV Broadcasting Station to a T-DMB Receiver

Kyoung Ho Bang, Young Cheol Park, and Jeongil Seo

ABSTRACT—We propose an efficient audio transcoding algorithm that can convert audio streams from terrestrial digital television broadcasting service stations to those for terrestrial digital multimedia broadcasting hand-held receivers. The proposed algorithm avoids the complicated psychoacoustic analysis by calculating the scalefactors of the bit-sliced arithmetic coding encoder directly from the signal-to-noise ratio parameters of the AC-3 decoder. The bit-allocation process is also simplified by cascading the nested distortion control loop. Through subjective evaluation, it is shown that the proposed algorithm provides comparable audio quality to tandem coding but it requires much smaller complexity.

Keywords—Audio coding, transcoding, AC-3, MPEG-4 BSAC, AAC, T-DTV, T-DMB, scalefactor.

I. Introduction

The emergence of streaming media players, coupled with the availability of powerful inexpensive laptop computers, PDAs, and hand-held mobile cellular phones, has created a domain for mobile multimedia applications. In Korea, the demand for mobile multimedia applications such as digital multimedia broadcasting (DMB) service has risen with the increasing popularity of these devices [1].

Recently, a number of international standards have been established based on different applications and technologies. The Dolby AC-3 audio compression standard [2] has been adopted for the DVD, terrestrial digital television (T-DTV), and high definition television (HDTV) in Korea and the United States. On the other hand, the new MPEG-4 audio standard [3] has a much broader and ambitious potential to support high and low bitrate multimedia applications for existing and future networks. MPEG-4 bit-sliced arithmetic coding (BSAC), in particular, has been selected as a Korean terrestrial digital multimedia broadcasting (T-DMB) [4] audio standard.

While these standards can operate for a spectrum of applications, each is optimized for a certain class of application. Sometimes, it is not economical or feasible to make any single media server or terminal to support all kinds of encoding and decoding. Transcoding techniques that convert one format to another, preferably in the compressed domain, can solve the problem of inter-standard operability.

In this letter, we present a new audio transcoding algorithm between an AC-3 decoder and a BSAC encoder. The algorithm proposed in this letter is to convert T-DTV audio streams into T-DMB audio streams with minimal complexity.

II. Proposed Audio Transcoding Algorithm

The block diagram of the proposed audio transcoding algorithm is shown in Fig. 1. The proposed algorithm employs the standard AC-3 decoder. The AC-3 bitstream is partially decoded to obtain signal-to-noise ratio (SNR) parameters. Later, fully decoded PCM samples are fed to the analysis filterbank of the BSAC encoder. But no psychoacoustic analysis module is used in the BSAC encoder. Instead, scalefactors are set from the SNR parameters of the AC-3 bitstream. The algorithm uses the same set of tools as the AC-3 decoder and the BSAC encoder.

1. Frame Synchronization

The objective of transcoding is to map the parameters of the source coder packed in a bitstream to those of a target coder.

Manuscript received Feb. 02, 2006; revised June 20, 2006.

Kyoung Ho Bang (phone: +82 2 2123 4534, email: euphony@cyclon.yonsei.ac.kr) is with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea.

Young Cheol Park (email: young00@dragon.yonsei.ac.kr) is with the Division. of Computer & Communications Engineering, Yonsei University, Wonju, Korea.

Jeongil Seo (email: seoji@etri.re.kr) is with Radio & Broadcasting Research Division, ETRI, Daejeon, Korea.

One important condition for transcoding is that the overall process should be performed in real-time based on synchronized source/target audio frames to make it useful for applications such as broadcasting and on-line streaming. The frame synchronization between the source and target coders must be achieved in such a way that $\Phi_t(i)=\Im_{ik}\{\Phi_s(k)\}$ where $\Phi_t(i)$ and $\Phi_s(k)$ denote parameter sets of source and target coders at the i-th and k-th frames, respectively.

In an AC-3 decoder, an audio frame comprises six 256-sample subblocks while BSAC employs 1,024-sample frames. Therefore, the frame synchronization is conveniently achieved by the following settings: $i=3m+l$; $k=2m+n$; $l=0, 1, 2$; $n=0, 1$; and $m=0, 1, 2$. Thus, two consecutive AC-3 frames are converted to three BSAC audio frames. Figure 2 shows the synchronization between AC-3 and BSAC frames.

If the block switch flag of at least one AC-3 audio subblock is 1, the corresponding BSAC frame is set to perform the block switching specified in the standard [3].

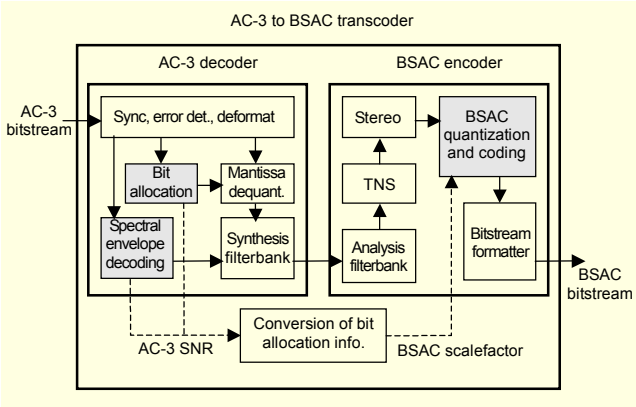

Fig. 1. Proposed AC-3 to BSAC audio transcoding algorithm.

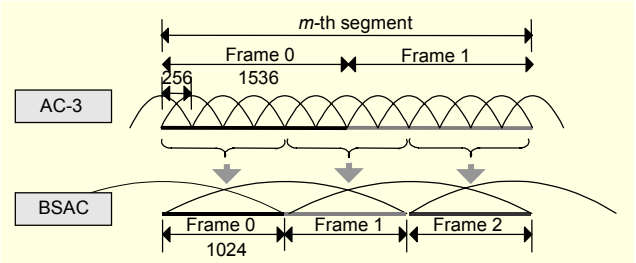

Fig. 2. Frame synchronization between the AC-3 decoder and the BSAC encoder.

2. Fast Bit Allocation

In the BSAC encoder, the modified discrete cosine transform (MDCT) coefficients are quantized and coded within two nested loops. In the inner (bitrate control) loop, the step size of the global non-uniform quantizer is adjusted not to exceed the number of bits available for one audio frame. In the outer (distortion control) loop, the step size is evaluated with respect to the psychoacoustic demands imposed by masking conditions. This is done in an analysis-by-synthesis procedure, which compares the actual quantized error to the previously calculated masking threshold and accordingly adapts the scalefactor. In the nested loop architecture, the inner loop is repeatedly run for each iteration of the outer loop, which results in significant computational loads. This problem can be solved by cascading the nested loops using a prediction model. Several methods have been suggested to implement this idea [5], [6].

In this letter, we use the scalefactor prediction model in [5], where the scalefactors are predicted using a statistical mode as

$$scf_b = \frac{8}{3}\log_2 \alpha_b - \frac{4}{3}\log_2 |X_b| + \frac{16}{3}\log_2 \frac{3}{4} + \varepsilon_b, \qquad (1)$$

where α_b is the masking threshold of scalefactor band b, X_b is the sum of magnitudes of the MDCT coefficients in band b, and ε_b is a relaxation constant controlling the shape of the quantization noise. By statistically predicting the scalefactors before the bit-allocation process, the distortion control loop is run just once before the rate control loop. This algorithm can save a significant amount of computational cost for the bit-allocation process when compared to the psychoacoustic model specified in MPEG audio because it requires the rate control loop to be run repeatedly for each iteration of the distortion control loop. Table 1 summarizes the computational gain obtainable by using the algorithm in [5].

Table 1. Complexity comparison between the nested and cascaded bit allocation loops.

	ISO	Proposed
Architecture	Dist. contr. loop N MIPS, J times Rate contr. loop M MIPS, K times	Dist. contr. loop N MIPS, 1 time Rate contr. loop M MIPS, K times
Complexity	$(N+M\times K)\times J$	$N + M\times K$

3. Parameter Conversion

To exploit the algorithm in (1), we need the MDCT coefficients X_b and the masking threshold α_b. The MDCT coefficients are directly obtained from the dequantization process of the AC-3 decoder. However, the masking threshold

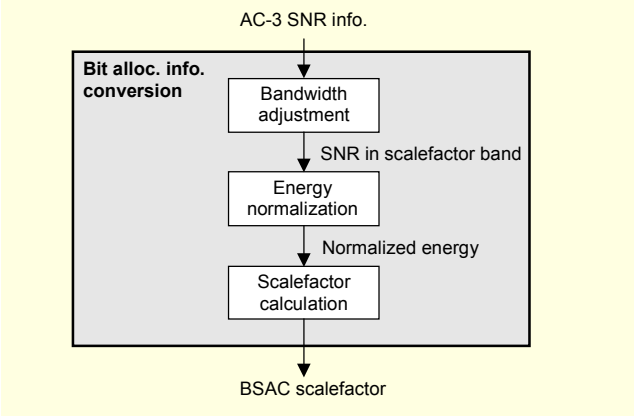

Fig. 3. Scalefactor calculation for the BSAC encoder.

is available only after the complicated psychoacoustic analysis. To avoid the psychoacoustic analysis we propose a method of converting the SNR parameters of the AC-3 decoder to the masking threshold of the BSAC encoder. Figure 3 shows the proposed parameter conversion process.

The AC-3 decoder computes the SNRs of 50 bands which have non-uniform bandwidths each with approximate 1/6 octave scale. From each SNR , we can compute the noise level of each band, being interpreted as the maximum allowable quantization noise as

$$\chi_c^u = PSD_c^u / SNR_c^u, \quad 1 \le c \le 50, \quad 0 \le u \le 3, \tag{2}$$

where c is the band number, u is the subblock index, and PSD_k is the power spectral density of band c. The subblock index u spans the range of 0 to 3, since, as we can see from Fig. 2, the frame synchronization obtains one BSAC frame from four AC-3 subblocks. Next, the band structure of the AC-3 decoder is reestablished onto 49 scalefactor bands of the BSAC encoder (bandwidth adjustment), and the noise level in the BSAC scalefactor band is obtained through parameter mapping:

$$\hat{\chi}_b^u = \chi_c^u, \quad 1 \le b \le 49, \quad 1 \le c \le 50, \quad 0 \le u \le 3. \tag{3}$$

Finally, the masking threshold of the BSAC scalefactor band b is estimated using the reestablished noise level as

$$\hat{\alpha}_b = \frac{\sum_{u=0}^{3} \hat{\chi}_b^u}{4 w_b}, \tag{4}$$

where w_b is the bandwidth of scalefactor band b. In (4) the reestablished noise level of scalefactor band b is averaged over four AC-3 subblocks, and it is normalized by the bandwidth of the corresponding scalefactor band (energy normalization). The estimated masking thresholds are then used to initialize the scalefactors of the BSAC encoder using (1) (scalefactor calculation). Figure 4 shows a typical set of scalefactors obtained using the transcoding process described by (2)-(4), and, for comparison, one obtained by the BSAC encoder in the tandem AC-3 decoder is also shown. Close resemblance between the two sets of scalefactors can be seen.

The computational advantages of the proposed algorithm are quite obvious. First of all, we can avoid the complicated psychoacoustic analysis by using the parameter mapping method in (4). Secondly, the exhaustive bit-allocation process of the BSAC encoder can be simplified because we use the scalefactor initialization in (1) instead of the nested loops. Another important point is that the bit-allocation process recommended by MPEG is not only computationally demanding but also largely varying in frame-by-frame computational load due to dynamic temporal and spectral variations of the input. It even fails at noise shaping, especially at low bit rates. Since the proposed algorithm is not associated with either the complicated psychoacoustic analysis or the nested bit-allocation loops, the frame-by-frame variation of computational complexity is also minimized.

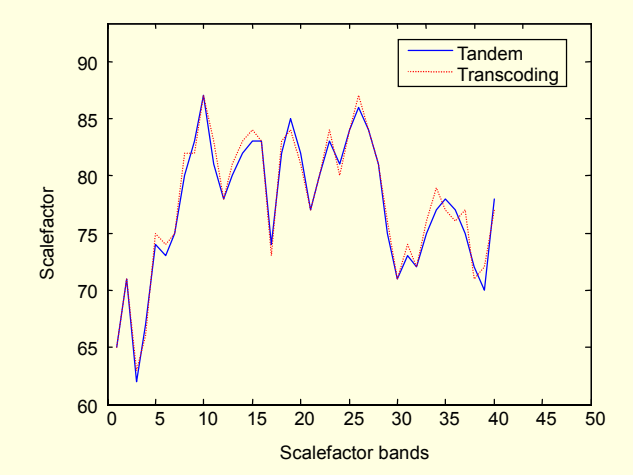

Fig. 4. Comparison of scalefactors determined through tandem and transcoding.

III. Performance Evaluation

To verify the efficiency of the proposed audio transcoding algorithm, we first examined the reduction factor of the computational load, and, secondly, we performed subjective tests.

The computational complexity of the proposed algorithm for one BSAC frame is summarized in Table 2, and is compared with the case of tandem coding in which the AC-3 decoder and BSAC encoder were simply cascaded. Since the psychoacoustic model and exhaustive bit-allocation iteration are not associated with the proposed algorithm, we could save almost 65% of the computational cost for the BSAC encoding.

For a subjective quality evaluation, we performed the double

Table 2. Computational complexity analysis.

		Tandem	Transcoding
AC-3 decoding			
	Overall process	26.53 (MOPS)	26.53 (MOPS)
BSAC encoding			
	T/F, TNS, etc.	5.23	5.23
	Psychoacoustic model	4.49	0
	Iteration loop	6.74	0.49
	Conversion logic	0	0.01
	Overall process	16.46 (MOPS)	5.73 (MOPS)

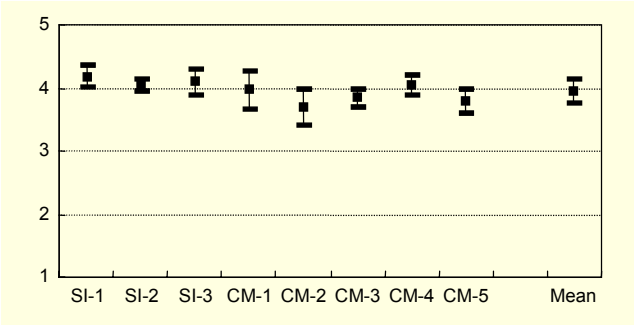

Fig. 5. Results of subjective tests for the tandem coder.

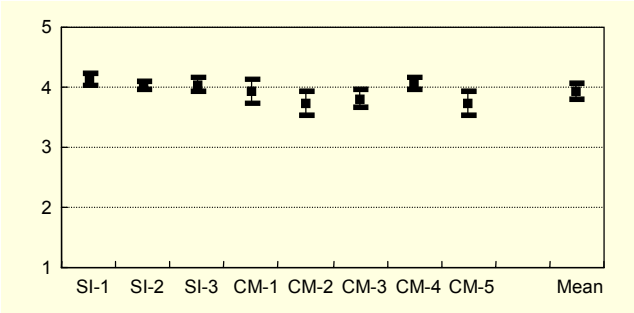

Fig. 6. Results of subjective tests for the transcoder.

blind triple stimulus with hidden reference tests described in ITU-R BS.1116 [7]. In the tests, twenty listeners who were trained and familiar with the test environment were involved. Each listener rated test materials heard through headphones using the 5-point-scale mean opinion score (MOS). The test materials consisted of three single instrumental sounds (SI) and five complex sound mixtures (CM). Each material was transcoded at 96 kbps stereo and compared with the sound obtained by tandem coding at the same bitrate. The mean scores of tandem and transcoding are 3.98 and 3.96, respectively. Details of the test results are shown in Figs. 5 and 6. According to the results, the proposed transcoding algorithm provides sound quality that is comparable to tandem coding. The listeners could not distinguish the materials synthesized using the proposed transcoding method from those using tandem coding. Therefore, the efficiency of the proposed algorithm is evident. It requires much less computational complexity while maintaining quality comparable to tandem coding.

IV. Conclusion

We presented an audio transcoding algorithm for bitstream conversion between an AC-3 decoder and a BSAC encoder. The algorithm calculates scalefactors for the BSAC encoder directly from the bit-allocation information of the AC-3 decoder. The proposed method allows us to save significant computation cost since it does not require the complicated psychoacoustic analysis and nested bit-allocation loops. Subjective tests revealed that the sound quality provided by the proposed algorithm is comparable to that of tandem coding.

References

[1] J. Seo, H. Moon, S. Beack, K Kang, and J-K Hong, "Multi-channel Audio Service in a Terrestrial-DMB System Using VSLI-Based Spatial Audio Coding," *ETRI J.*, vol.27, no.5, Oct. 2005, pp.635-638.

[2] C. Todd et al., "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," *AES 96th Convention*, 1994.

[3] ISO/IEC JTC1/SC29/WG11, *Information Technology – Coding of Audiovisual Objects Part 3: Audio*, FDIS 14496-3 1998.

[4] TTA Korea, *Digital Multimedia Broadcasting*, SG05.02-046, 2003.

[5] K.H. Bang, K.S. Lee, Y.C. Park, and D.H. Youn, "Fast Bit Allocation Method for MP3/AAC Encoders," *AES 118th Convention*, 2005.

[6] H.O. Oh, J.S. Kim, C.J. Song, Y.C. Park, and D.H. Youn, "Low Power MPEG/audio Encoders Using Simplified Psychoacoustic Model and Fast Bit Allocation," *IEEE Transaction on Consumer Electronics*, vol.47, Aug. 2001, pp.613-621.

[7] ITU-R, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multi-channel Sound Systems*, ITU-R Recommendation BS.1116, 1994.