AUDIO TRANSCODING ALGORITHM FOR MOBILE MULTIMEDIA APPLICATION

Kyoung Ho Bang, Young Cheol Park**, and Dae Hee Youn**

*DSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, KOREA
**Division of Computer and Telecommunications Eng., Yonsei University, Wonju-city, KOREA

E-mail: euphony@cyclon.yonsei.ac.kr

ABSTRACT

In this paper, we propose an audio transcoding algorithm for mobile multimedia applications. The algorithm is to provide high quality transcoded audio stream from a multimedia broadcasting media server system to mobile terminals. As a representative case, audio transcoder between T-DTV and T-DMB service is considered. While the Korean DTV audio standard adopted the Dolby AC-3, the Korean T-DMB service uses the MPEG-4 BSAC. The proposed algorithm reuses the bit allocation information of the AC-3 decoder in the process of BSAC encoding and simplifies the bit-allocation process with two independent loops instead of nested ones. We verified that our architecture could take the place of the nested loop with a significant simplification and no PAM. The proposed audio transcoding algorithm is suitable for the mobile multimedia applications due to its low complexity and minimum frame-by-frame complexity variations.

1. INTRODUCTION

The emergence of streaming media players, coupled with the availability of powerful inexpensive laptop computer, PDAs and hand-held mobile cellular phone has created a domain for mobile multimedia applications. In Korea, mobile multimedia applications such as Digital Multimedia Broadcasting (DMB) service arise with tremendous popularity. Unfortunately, a mobile multimedia terminal should be low-powered for battery operation and multimedia applications including real-time audio and video through narrow band mobile channels ask for huge processing amount in order to realize bandwidth compression/decompression of audio and video in real-time.

Recently, a number of international standards have been established based on different applications, technology and era. While these standards can operate for a spectrum of applications, each is optimized for a certain class of application. Sometimes, it is not economical and feasible to make any single media server or terminal to support all kinds of encoding and decoding. Therefore, transcoding techniques that convert one format to another, preferably in the compressed domain, are required to solve the problem of inter-standard operability.

Among audio standards, the Dolby AC-3 audio compression standard [1] is successfully adapted in the Digital Versatile Disc (DVD) and Terrestrial Digital Television (T-DTV) broadcast industries. Now a large variety of products based on the AC-3 standard are available on the market. On the other hand, the new MPEG-4 audio standard [2] has much broader and ambitious perspective to support high and low bitrate multimedia applications to existing and future networks. MPEG-4 Bit-Sliced Arithmetic Coding (BSAC) is selected as a Korean Terrestrial Digital Multimedia Broadcasting (T-DMB) [3] audio standard.

Many audio-visual programs have been produced for T-DTV broadcasting service for a long time. These programs have MPEG-2 encoded video stream and AC-3 encoded audio stream. Unfortunately we cannot reuse the audio stream of T-DTV program for T-DMB service directly because of adopting the different audio format, MPEG-4 BSAC in T-DMB service. For T-DMB service, a large amount of audio-visual programs should be produced newly or converted into BSAC format.

In this paper, we present a new audio transcoding algorithm between AC-3 to BSAC. Using this algorithm, we can convert the T-DTV audio stream into the T-DMB audio stream to have a minimum complexity.

2. MOBILE MULTIMEDIA SERVICES AND AUDIO CODING STANDARDS

2.1. AC-3 Audio Coding and T-DTV Service

The Dolby Laboratory developed AC-3 audio coding scheme. Dolby Digital (the audio standard used in film industry DVD, multimedia, HDTV), Dolby Surround Digital (the audio standard used in Home Theater System (HTS)), and Dolby Net (the audio standard used in the internet environment), all refer to the same kernel – the AC-3 audio coding technology. To feed these quite different demands of all these different areas, the target bitrates of AC-3 ranges from 32kbps to 640kbps.

AC-3 is a perceptual audio coder (PAC), that is, AC-3 uses human psychoacoustic features to mask the inaudible audio signals, so the bits are saved for representing really important signals. Up to 5 full-bandwidth channels and one subwoofer channel (cutoff at 120Hz) can be contained in an AC-3 bitstream (this is the so-called 5.1 channels). The time-frequency transform used in AC-3 is the Analysis/Synthesis Filter Bank with Time Domain Aliasing Cancellation. A parametric bit allocation process is applied in AC-3, so flexibility and efficiency can be achieved at the same time. With sophisticated psychoacoustic model and the decorrelation between neighboring frequencies and different channels, AC-3 provides low-bandwidth but high quality perceptual audio sound.

AC-3 audio coding has been adopted by the Advanced Television System Committee (ATSC) as the audio service standard for High Definition Television (HDTV) in the United States. It has also found applications in consumer media (laserdisc, digital versatile disc) and direct satellite broadcast.

2.2. BSAC Audio Coding and T-DMB Service

BSAC is a part of the newest audio coding standard from ISO/IEC known as ISO/IEC 14496-3 (MPEG-4/Audio) [2]. BSAC is mainly based on MPEG-4 AAC and uses most of its tools. BSAC improves upon MPEG-4 AAC by offering a fine grain scalability and error resilience. Its compression rate is comparable to the AAC Main profile.

MPEG-4 BSAC uses the same set of tools as MPEG-4 AAC including 1024/128 point Modified Discrete Cosine Transform (MDCT), Temporal Noise Shaping (TNS), Long Term Prediction (LTP), Noiseless coding, and so on. The main features of BSAC are as follows.

- Fine grain scalability
- Efficient audio compression
- Error resilience

The scalable coder achieves audio coding at different bitrates and qualities by processing the bitstream in an ordered set of layers. The base layer is the smallest sub-set of the bitstream that can be decoded independently to generate the audio outputs. The remaining bitstream is organized into a number of enhancement layers such that each enhancement layer improves the audio quality. BSAC supports a wide range of bitrates from the low bitrate stream of 16kbps per channel (kbps/ch) at the base layer to the higher bitrate stream of 64kbps/ch at the top layer. BSAC offers a fine-grain scalability at 1kbps/ch. This fine grain scalability of BSAC comes from the bit-sliced coding technique. In bit-sliced arithmetic coding, the quantized spectral values are first grouped by the frequency bands. Then, the bits in each group are processed in slices, i.e., “bit-sliced”, in the order from MSB (most significant bits) to LSB (least significant bits). The most significant bits across the groups form the first bit-slice. This bit-slice is fed into the noiseless coding part and then transmitted in the base layer. The next significant bits form the second bit-slice, which is processed and transmitted in the first enhancement layer. This process continues until the least significant bits are processed and transmitted in the top enhancement layer. Each enhancement layer adds 1kbps/ch of bitstream and thus provides fine grain scalability in BSAC.

The bit-slices that are fed into the noiseless coding part go through an entropy coding process. In BSAC, the entropy coding is carried out by an arithmetic coder. In AAC, a Huffman coder is used instead. The arithmetic coder in BSAC enhances the coding efficiency of the bit-sliced audio streams.

The error resilience feature of BSAC is implemented by Segmented Binary Arithmetic (SBA) coding. In SBA, multiple layers of audio streams are grouped again into segments. Any error propagation is constrained to a single segment in BSAC by reinitializing the arithmetic coder after every N^{th} enhancement layer.

The audio service in DMB should support the standardized stereo audio broadcasting at the sampling rates of 24, 44.1, or 48kHz. The service should provide CD-quality audio for the audio-only broadcasting and better than the analog FM radio for the audio accompanying the video. The maximum bitrate for the audio data in stereo is set to 128kbps.

DMB employs the MPEG-4 BSAC standard’s LC profile that is suitable for mobile applications. BSAC allows adaptive bitrate control, a smaller initial buffer, and a seamless play of digital audio.

3. PROPOSED AUDIO TRANSCODING ALGORITHM

3.1. AC-3 to BSAC Transcoder

The block diagram of the propose AC-3 to BSAC audio transcoder is shown in Figure 1. The overall system has no psychoacoustic analysis module in BSAC encoder and additional logic that converts SNR information of AC-3 decoder to the scalefactors for BSAC encoder. Beside these two differences, transcoder uses the same set of tools as general AC-3 decoder and BSAC encoder.

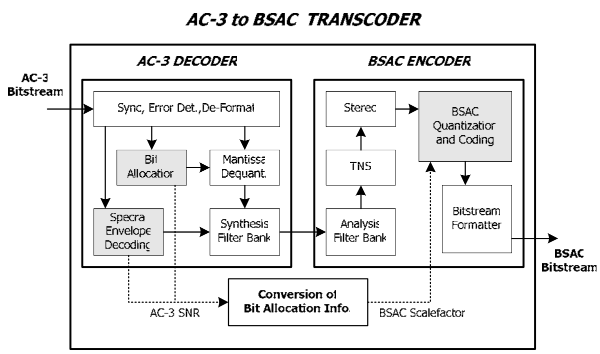

Figure 1. Proposed AC-3 to BSAC audio transcoder

3.2. Reusing Bit Allocation Information

The quantizer bit allocation problem is to determine the individual rates of a finite collection of quantizers so as to minimize the sum of their distortions, subject to a constraint on the sum of their rates. Both AC-3 and MPEG-4 BSAC audio coding standard are based on the perceptual audio coding technology using masking property of the human auditory system. A numerical expression of the masking property in frequency domain is the masking threshold calculated from the psychoacoustic analysis to require amazing computational complexity and memory resources.

To lay down this burden, we design a special logic that converts the bit allocation information of the AC-3 decoder into the scalefactors of the BSAC encoder without psychoacoustic analysis process. Figure 2 shows the conversion module of bit allocation information.

Bandwidth Adjustment re-establishes the 64 subbands in AC-3 decoder by the 49 scalefactor bands of the BSAC encoder. Energy Normalization process builds up the SNRs of the four AC-3 frames.

First, we calculate the arithmetic mean of the SNRs for four successive AC-3 frames using Equation (1).

$$A[i] = \frac{B_n[i] + B_{n+1}[i] + B_{n+2}[i] + B_{n+3}[i]}{4} \qquad (1)$$

$$B_n[i] = PSD_n[i] \,/\, SNR_n[i]$$

where $B_m[i]$ is the masking threshold converted from the SNR of the i^{th} subband in the m^{th} frame.

And then the means in each scalefactor band divided by the bandwidth of the corresponding band:

$$A_N[i] = \frac{A[i]}{k} \qquad (2)$$

where k is the bandwidth of the i^{th} scalefactor band.

Scalefactor Calculation [4] computes the scalefactors of the BSAC encoder from normalized threshold ($A_N[i]$), the summation of MDCT coefficients of BSAC encoder in the k^{th} scalefactor band (X_k), system constants (c_k):

$$scf \le \frac{8}{3}\log_2 A_N - \frac{4}{3}\log_2\left|X_k\right| + \frac{16}{3}\log_2\frac{3}{4} + c_k \qquad (3)$$

Through this conversion logic, outer iteration loop can be removed and inner iteration loop simplified in BSAC encoding process. Moreover, psychoacoustic analysis need not computed due to the presence of bit allocation information.

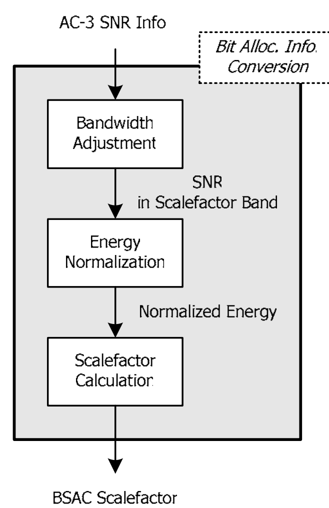

Figure 2. Conversion of Bit Allocation Information

4. PERFORMANCE EVALUATION

To verify the efficient audio transcoding algorithm, two verifications were applied to the suggested algorithm. First verification was to examine the reduction factor of the processor load, and second one was to verify the performance of subjective quality test.

Table 1. Computational Complexity Analysis

	Tandem	Transcoding
AC-3 Decoding		
Overall process	**100%**	**100%**
BSAC Encoding		
T/F, TNS, etc.	30%	30%
PAM	30%	0%
Iteration Loop	40%	3%
Conversion Logic	0%	2%
Overall process	**100%**	**35%**

The computational complexity of one BSAC frame is summarized in Table 1. In this table, proposed audio transcoding algorithm combines no psychoacoustic model, conversion process and simplified iteration loop with speed up scheme. It is clear that proposed transcoding algorithm take advantage of the computational complexity.

For a subjective quality evaluation, we performed the double blind triple stimulus with hidden reference test described in ITU-R BS.1116 [5]. In the test twenty listeners who were trained and familiar with test environment were involved. Each listener rated test material using the ITU-R 5 point scale. The test material was heard through headphone. Test material was composed of three single instrumental sounds (SI) and five complex sound mixtures (CM), and each material was made by tandem coding and transcoding between AC-3 to MPEG-4 BSAC at 96kbps stereo using the reference tandem coder and the proposed transcoder, respectively. Table 2 shows the test results.

Table 2. Subjective Sound Quality Test

	Tandem	Transcoding
SI-1	4.18	4.13
SI-2	4.05	4.03
SI-3	4.10	4.05
CM-1	3.97	3.92
CM-2	3.70	3.75
CM-3	3.85	3.81
CM-4	4.05	4.06
CM-5	3.79	3.74
Average	3.96	3.94

According to the results in Table 2, the proposed transcoding algorithm is capable of providing resembling sound quality than tandem coding. The listeners cannot distinguish the material using the proposed transcoding method from those using the tandem coder practically.

Thus, the efficiency of the proposed algorithm is evident, since it requires much less computational complexity but performs better than the tandem coding.

5. CONCLUSIONS

We presented an audio transcoding algorithm between AC-3 to BSAC audio coding. It is based on the reusing bit allocation information. The method was obtained by converting the SNR of the AC-3 decoder into the scalefactors of the BSAC encoder. The proposed method mainly allows us to save significant computation time.

The psychoacoustic analysis and bit allocation method recommended by MPEG requires a large amount of complexity and a large variation of that due to the nested loop architecture, and it often gives up the noise shaping due to the escape condition of the nested loop at very low bit rates. We reused the bit allocation information and divided the nested loop into two independent loops and verified that our architecture could take the place of the nested loop with a significant simplification and no PAM. In order to reduce the variation of the complexity of loop architecture, we replaced the distortion control loop by the scale factor initialization algorithm. Thus, the bit allocation process can be replaced by a loop-free process, and it can reduce the variance of the complexity and moreover a large amount of complexity. The proposed audio transcoding algorithm is efficient for mobile multimedia application.

6. REFERENCES

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

[2] ISO/IEC JTC1/SC29/WG11 No. 2203 IS 14496-3 (Information Technology – Coding of Audiovisual Objects Part 3. Audio),” 1998.

[3] “Digital Multimedia Broadcasting,” *Telecommunication Technology Association in Korea*, 2003SG05.02-046, 2003.

[4] Kyoung Ho Bang, Keun Sup Lee, Young Cheol Park, and Dae Hee Youn, “Fast bit allocation method for MP3/AAC Encoders,” AES 118th Convention, 2005.

[5] ITU-R, Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems, ITU-R Recommendation BS. 1116, 1994.