Flexiable Audio System for Multipurpose

Choong Sang Cho, Je Woo Kim, Hwa Seon Shin and Byeong Ho Choi
SoC Research Center, Korea Electronics Technology Institute
68 Yatap-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea
E-mail: {ideafisher, jwkim, L544, bhchoi}@keti.re.kr

Abstract: In this paper, we design a flexiable audio system for multipurpose with MPEG surround, MPEG-2 AAC and MPEG-4 ALS. It has the three processing levels. At the first level, the multi-channel audio is down mixed to stereo signal with spatial parameter. The down mixed signal is encoded with the MPEG-2 AAC. At the second level, the multi-channel is down mixed as the first level and the stereo signal is encoded with the MPEG-4 ALS. At the third level, the multi-channel is encoded using the MPEG-4 ALS without data loss. The performance of the proposed audio system is evaluated by comparing the compression ratio at each level when we use MPEG-4 ALS conformance test data. The proposed audio system has the three levels that have the different compression ratio and data loss. Therefore, the proposed audio system can be used for multipurpose.

1. Introduction

With the popular use of the multi-media products, the audio codec has been widely used for the communication and the storage products. The development of audio system for storage products and communication is differently focused on the usage and the purposes. The audio system for storage product focuses on the audio quality than compression ratio, when the products have enough space to save the encoded audio data. While the audio system for communication is developed to provide the high compression ratio due to the real-time audio services. Therefore many kind of audio codec are used based on the purpose. Also the audio codec have to cover the multi-channel audio to give the high quality audio effects in this time. The requirement of high quality audio increases the size of audio data and complexity. In this paper, we propose an audio system that can be used for multipurpose. The proposed audio system provides the multi-channel audio for the high audio quality, also it has the multi level for audio quality and compression ratio. Therefore, the proposed audio system can be used for the storage product and communication, etc. Following this introduction, the overall structures of the used audio coders are described in Section 2. In Section 3, we address the proposed audio system for multipurpose. In section 4, the proposed audio system is evaluated and, we conclude this paper in Section 5.

2. Used Audio Coders for Proposed Audio System

In this section, we decribe the audio coders that are used to compose the proposed audio system.

2. 1 MPEG Surround Audio Coder

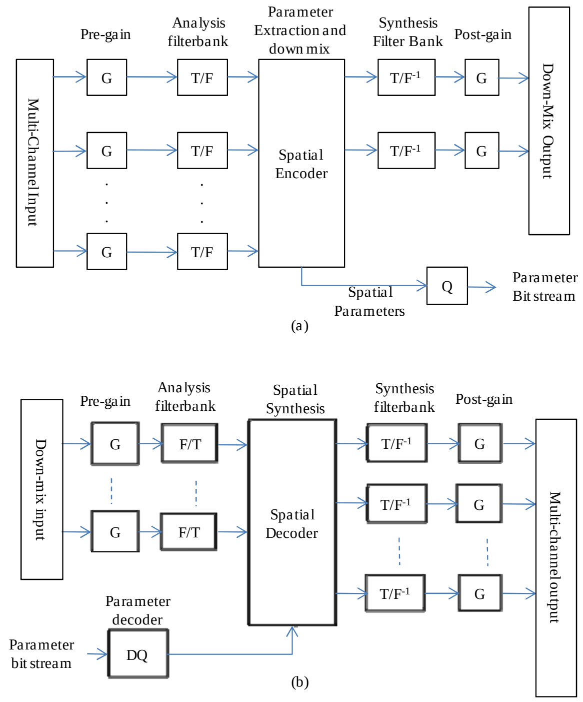

Figure 1. Structure of MPEG surround; (a) Encoder structure, (b) Decoder structure.

MPEG surroud developed to provide the high quality multi channel audio and high compression by down mixing the multi channel to stereo or mono [1]. The down mixed signal is achived by extracting the spatial image information based on SAC(Spatial Audio Coding). The Structure of MPEG surround is shown as Figure 1 [1][2]. To down mix the multi channel audio, the amplitude of multi-channel audio signal is scaled by pre-gain. The scaled multi-channel audio signal is analyzed at the analysis filter bank. The analysis filter bank is consist of 32 or 64 band QMF(Quadrature Mirror Filterbanks) [1]. The band of QMF which is composed based on auditory sense of human has high analysis resolution at low frequency, while it has low analysis resolution at high frequency. To improve the resolution of low frequency, the MPEG surrournd has the hybrid QMF. The three low bands of QMF are analyzed more detailly by hybird QMF. The lowest QMF band is seperated into eight bands, the second and third QMF bands are seperated into two bands. The output of hybrid QMF is used to calculatete the power of each parameter band.
In spatial encoder block, the spatial image parameters and down mixed singal are obtained. For the spatial image parameter, the CLD(Channel Level Difference) and ICC(Interaural Level Difference) are obtained. The CLD expresses the power ratio of two channels in logarithm and the ICC expresses the correlation of two channels [2]. The

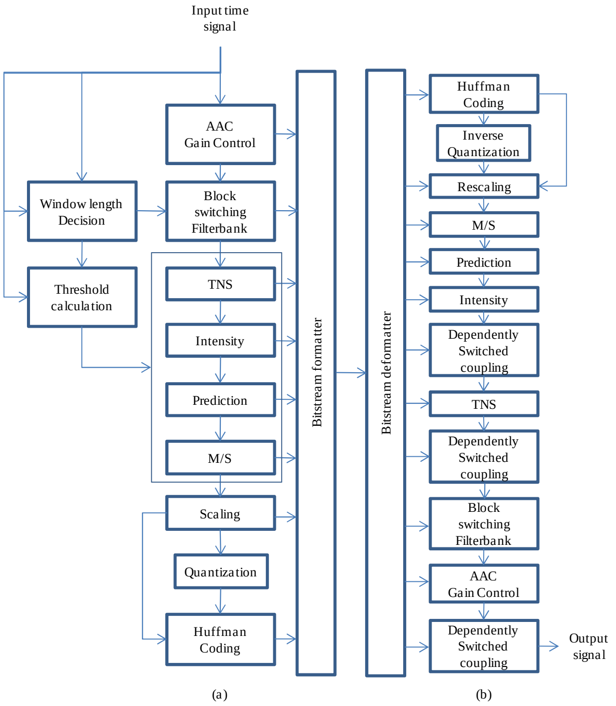

Figure 2. Structure of MPEG-2 AAC; (a) Encoder structure, (b) Decoder structure.

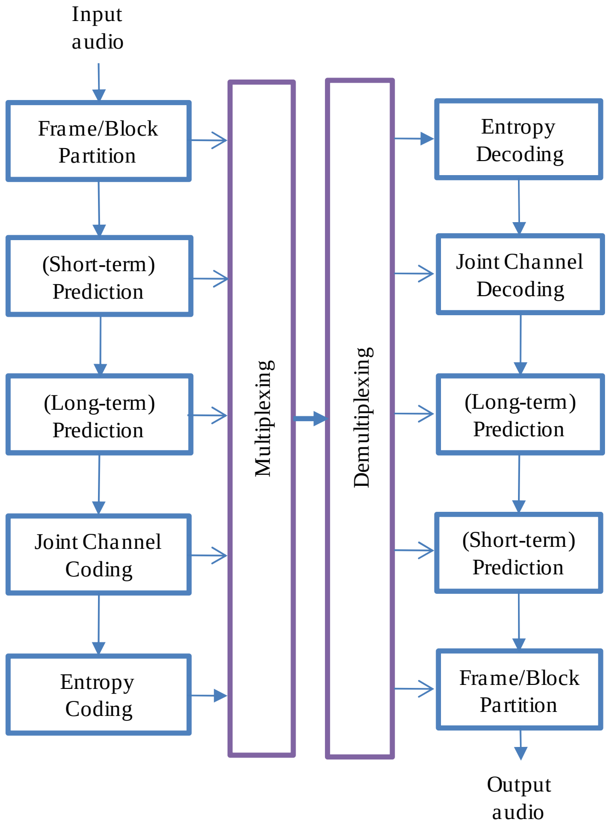

Figure 3. Structure of MPEG-2 AAC; (a) Encoder structure, (b) Decoder structure.

obtatined spatial parameters at parameter bands are encoded by using the Huffman coding. To down mix the multichannel audio signal into stereo or mono, the two or three channels are coupled and the coupled channels are added after applying the weight to each channel.

The down mixed signal is converted into audio signal by using the synthesis filter bank that is consist of inverse QMF. Therefor, the multichannel audio signal is converted into spatial parameters and down mixed audio signal by using the MPEG surround encoder. The MPEG surround decoder has the inverse procedure of MPEG surround encoder as shown Figure 1.

2. 2 MPEG-2 AAC Audio Coder

MPEG-2 AAC(Advanced Audio Coding) has been used for broadcasting service because it has the high audio quality at low bitrate. The structure of MPEG-2 AAC is shown as Figure 2 [3][4]. The audio input are converted into frequency domain by using the filterbank that is MDCT(Modified Discrete Cosine Transform) . In filterbank, different kind of block is switched to reduce the pre-echo. But the block switching can't perfectly remove the pre-echo, therefore, the TNS(Temporal Noise Shaping) is applied to remove the pre-echo by using the duality of time and frequency [3][4].

The intensity block is used to compress the high frequency region more efficiently, also the prediction and M/S(Mid-Side) are applied to reduce the amplitude of the data. The output of MS stereo is scaled and quantized. The quantized data are encoded by using the Huffman coding without lossless. The decoder of MPEG-2 AAS is operated by the inverse procedure of the encoder.

2. 3 MPEG-4 Audio Lossless Coder

MPEG-4 ALS is a lossless audio coding that ensures the perfect reconstruction of the original audio data that were input to the encoder. MPEG-4 ALS supports the high sampling frequency up to 192 kHz and the bit resolution for data has up to 32 bit PCM and IEEE-754 32 bit floating point. Also MPEG-4 ALS supports the many channels up to 65536. The MPEG-4 ALS encoder and decoder have the structure as shown Figure 3 [5].

The MPEG-4 ALS separates a frame into several blocks in frame/block partition block. The block is unit to process entropy coding to reduce the size of encoded data [5]. The input speech of MPEG-4 ALS is used to calculate the LPCs(Linear Prediction Coefficients) in short-term prediction. To obtain LPCs, the MPEG-4 ALS calculates the autocorrelations of input speech, and then the autocorrelations are applied to Levison Durbin Algorithm [5][6]. The LPCs is used to predict the current sample of a time-discrete signal $x(n)$ from the previous samples $x(n-k)$. The prediction is accomplished by using the FIR filter as Eq1 where the filter uses LPCs as filter coefficient

$$\hat{x}(n) = -\sum_{i=1}^{K} \gamma(i) x(n-i) \qquad (1)$$

In here, $\hat{x}(n)$ is the predicted sample, $\gamma(k)$ is LPCs and K is the order of FIR

The residual for transmission is obtained by subtracting the predicted sample from the original sample. The long-term prediction block is used to cover the distance correlation based on pitch period. By applying the residual to long-term prediction block, the amplitude of residual is decreased [5][6]. The MPEG-4 ALS uses the joint channel coding block which has the two kind methods to reduce the required bit size for encoding. When the stereo input channels have high correlation, the difference between two channels is calculated and encoded. Another method to improve the compression performance in joint channel coding block is multi channel coding. The each channel value is subtracted from reference channel with applying weight in 3 taps and 6 taps filtering. The reference channel and subtracted valued are transmitted. The obtained residual from joint channel coding block is encoded in entropy coding block. The encoding block uses the two encoding method [6]. The first one is rice coding. The rice coding calculates the rice parameter from average of residual of a block. The residual of a block is encoded with each rice parameter. The second one is BGMC (Block Gilbert-Moore Codes). In BGMC coding, residual is separated into two categories by using the distribution. The tail region is encoded by rice coding and the central region is encoded by Block Gilbert-Moore codes and direct fixed-length codes.

3. Proposed Audio System for Multipurpose

In this paper, the flexible audio system is proposed that is suitable for multipurpose and multi-channel audio. The proposed multipurpose audio system has the structure as shown Figure 4. To be used for multipurpose, the proposed audio system has the three compression ratio level and audio quality levels as Figure 4(a). In the proposed audio system, the audio processing level is selected based on the used purpose by enabling the required blocks for each level.

When the highest compression ratio is required to service the multi-channel audio at the real time audio broadcasting and when the storage space for multi channel audio is not enough at the portable multi-media devices, the first level of the proposed audio system is selected and it is the highest compression ratio level in the proposed audio system. The first level is consists of MPEG surround encoder and MPEG-2 AAC encoder that are lossy audio coders. The multi-channel audio is down mixed into stereo audio and spatial image parameter by using the MPEG surround encoder. And the down mixed audio is encoded by MPEG-2 AAC encoder, because the MPEG-2 AAC provides the good audio quality at the low bitrate. The size of multi-channel audio data is suddenly reduced by MPEG surround, also the size of down mixed audio is more reduced by using the MPEG-2 AAC. The spatial image parameter is packetized without any coding. The first level has the loss of the audio data by the two lossy audio coders, but it has the highest compression ratio in the proposed audio system.

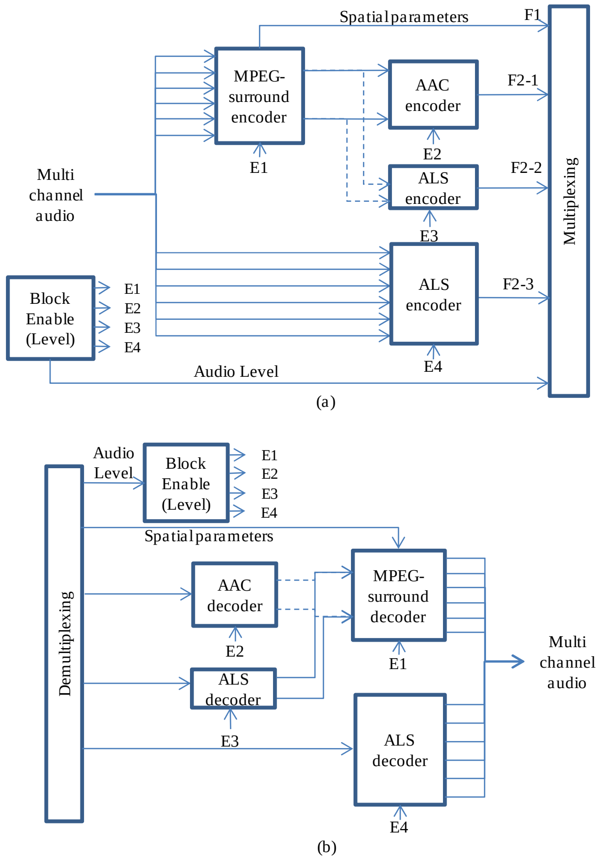

Figure 4. Structure of proposed audio system for multipurpose; (a) Encoder structure, (b) Decoder structure.

The second level of the proposed audio is consists of MPEG surround encoder and MPEG ALS encoder. The second level can be selected, when the priority of compression ratio is decreased and the priority of audio quality is increased. In second level, the multi channel audio is converted into down mixed audio and spatial image parameter as the first level. The down mixed audio signal is compressed without losses by using MPEG ALS encoder. By using the lossless coding, the audio quality is improved, while the compression ratio is deceased, when it is compared with the first level. But the second level compression ratio is high as much as the first level due to the MPEG surround. Therefore, the second level can be used to service the high quality multi-channel audio broadcasting, also it can be applied to high compression and high quality multi-media products.

The third level is consists of MPEG ALS encoder. It is selected, when the audio quality is the best important component and the storage or the bandwidth for transmission is enough. The third level of the proposed audio system compresses the input audio of multi-channels by using MPEG ALS without any losses. The third level provides the highest audio quality in the system, while the compression ratio is the lowest in the system. Therefore, the third level can be applied to storage products that have enough space and high quality audio broadcasting service when it has a wide bandwidth for transmission.

The decoding structure of the proposed audio system is shown as Figure 4(b). The audio level is used to enable the required block on each level. When the first level is selected, the MPEG-2 AAC decoder block and MPEG-surround block are enabled. The bitstream of MPEG-2 AAC and spatial image parameter are unpacked for the first level. The down mixed audio is obtained by using the MPEG-2 AAC decoder. The multi-channel audio is decoded by using the MPEG-surround decoder with the down mixed audio and spatial image parameter. When the second level is selected, the MPEG ALS decoder block and MPEG-surround block are enabled, the down mixed audio is obtained by using MPEG ALS decoder and multi-channel audio is decoded as the first level. As the third level is selected, the MPEG ALS is only enabled and the multi-channel audio is decoded without any loss.

4. Performance Evaluation

The performance of the proposed audio system was evaluated by the compression ratio of the multi-channel audio. To compare the compression ratio as Eq 2, we used the MPEG-4 ALS conformance test file that is 96 kHz, 6 channels, 24 bit data resolution [7].

$$Compression\ ratio = \frac{Compressed\ \ file\ size}{Original\ \ file\ size} \times 100 \quad (2)$$

Table 1. Compression ratio of each level in the proposed audio system.

File Type (Level)	File Size (Byte)	Compression Ratio(%)
Original File	23,922,602	-
First Level	264,300	1.10
Second Level	1,236,076	5.16
Third Level	8,912,425	37.25

The Table 1 shows the size of compressed file and compression ratio when the original file is processed by the proposed audio system. When the original file was compressed on the first level, the size of original file was decreased into 264,300 Byte, and the compression ratio was 1.10 %. When we applied the second level of the proposed audio system, it was compressed into 1,236,076 Byte and the compression ratio was 5.16%. In the third level, the original file was compressed to 8,912,425 byte and the compression ratio was 37.25%.

5. Conclusion

In this paper, we proposed the flexible audio system for multipurpose. The proposed audio system is consists of MPEG surround, MPEG-4 ALS and MPEG-2 AAC. The proposed audio system has the three kind of processing levels. The first level of the system was composed of MPEG surround and MPEG-2 AAC. It has the highest compression ratio in the proposed audio system. But the first level has the loss of audio data by using the MPEG surround and MPEG-2 AAC. The second level is consists of MPEG surround and MPEG-4 ALS. It has less compression ratio than the first level, but it has the only loss of audio data by using the MPEG-surround. In the third level, the MPEG-4 ALS was only used. It has the lowest compression ratio in the proposed audio system. But the original audio data is reproduced without any loss.

References

[1] ISO/IEC 23003-1, *Information technology – MPEG audio technologies*, part 1: MPEG surround, 2006.

[2] J. Breebaart, C. Faller, "*Spatial Audio Processing MPEG Surround and Other Application*," John Wiley & Sons, 2007.

[3] ISO/IEC 13818-7, *Information technology –Coding of audio-visual objects*, MPEG-2 AAC, 2005.

[4] M. Bosi, R.E. Goldberg, "*Introduction To Digital Audio Coding and Standards*," Kluwer Academic Publishers, 2003.

[5] T. Liebchen, Y Reznik, T. Moriya, D.T. Yang, "MPEG-4 Audio Lossless Coding," in *Proc. of 116th AES Convention*, Preprint No. 6047, May 2004.

[6] ISO/IEC 14496-3:2005/Amd 2:2006. *Audio Lossless Coding (ALS), new audio profiles and BSAC extensions*, 2006.

[7] ISo/IEC 14496-4:2004/FDAM 19:2007(E), *Information technology – Coding of audio –visual objects – Part4: Conformance testing*, Amendment 19: Audio lossless coding, 2007.