THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.28, No.1E 2009. 3. pp. 29~34

Dual-Domain Connection Scheme for HE-AAC and MPEG Surround

Hee-Suk Pang*
*Dept. of Electronics Engineering, Sejong University
(Received February 20 2009; accepted March 23 2009)

Abstract

MPEG4 High Efficiency Advanced Audio Coding (HE-AAC) and MPEG Surround are one of the most efficient combinations for low bit rate multi-channel audio coding. Based on the fact that these two codecs have identical quadrature mirror filter (QMF) analysis and synthesis structures, we propose a dual-domain connection scheme for the codecs. Specifically two time-domain connection methods are analyzed and compared to the QMF subband-domain connection method. Experimental results show that both the time-domain connection methods cause no subjective sound quality degradation compared to the QMF subband-domain connection method, which verifies that one can select either of them depending on application scenarios.

Keywords: *High Efficiency Advanced Audio Coding, MPEG Surround, Quadrature mirror filter, Dual-domain connection*

I. Introduction

Recently MPEG Surround was standardized to achieve both a low bit rate and high sound quality for multi-channel audio coding [1]. An MPEG Surround encoder downmixes original multi-channel audio signals and extracts spatial parameters composed of channel level differences (CLDs), inter-channel correlation /coherences (ICC), and channel prediction coefficients (CPCs). Then an MPEG Surround decoder reconstructs multi-channel audio signals based on the downmix signal and spatial parameters. Due to its low bit rate, MPEG Surround is suitable for streaming services such as multi-channel digital audio broadcasting, which can be played by home and car multi-channel playback systems. For technical description of MPEG Surround, refer to [2-4].

Corresponding author: Hee-Suk Pang (hspang@sejong.ac.kr)
Dept. of Electronics Engineering, Sejong University, Gunja-Dong 98, Gwangjin-Gu, Seoul, Korea

MPEG Surround can be used with any downmix codec that is used to encode and decode the downmix signal. From the viewpoint of a bit rate, HE-AAC [5] or aacPlus [6] is one of the most efficient downmix codecs. Since HE-AAC utilizes the spectral band replication (SBR) tool to generate high frequency components from the low frequency components, it reproduces good quality audio at the bit rate of 32-64 kbps [7]. As a result, HE-AAC and MPEG Surround can be used together to reproduce good quality 5.1-channel sounds at the total bit rate of 64 kbps or less. Furthermore, HE-AAC is currently used for digital multimedia broadcasting (DMB) [8], which means that the current DMB systems can be upgraded from stereo to 5.1-channel by adopting the MPEG Surround technology.

In the meanwhile, HE-AAC and MPEG Surround are both based on QMF analysis and synthesis structures. Especially at sampling frequencies of 32 kHz, 44.1 kHz, and 48 kHz, which are the most common cases in audio coding, HE-AAC and MPEG Surround use identical QMF structures [1, 5]. Consequently, QMF subband samples can be directly connected between them without QMF synthesis in HE-AAC and QMF analysis in MPEG Surround, which is efficient for both computational complexity and memory requirement. On the other hand, the time-domain connection is still necessary in case that only PCM inputs and outputs are available between them. For example, HE-AAC and MPEG Surround can be implemented on different chipsets where only PCM samples are transmittable and receivable. Based on this practical problem, the idea of dual-domain connection for HE-AAC and MPEG Surround was presented in [9] and adopted for the MPEG Surround standard [1]. Whereas the MPEG Surround standard specifies that both the QMF subband-domain and time-domain connections should be supported between HE-AAC and MPEG Surround [1], the detailed technical description on the scheme has not been sufficiently made in the literature so far. In this paper, we give a detailed description and experimental results of the dual-domain connection scheme for the two codecs. In the following sections, we assume that the sampling frequency of HE-AAC is one of 32 kHz, 44.1 kHz, and 48 kHz.

II. The Dual-Domain Connection Scheme

The MPEG Surround decoding process is as follows [1]. A time-domain signal is first analyzed by QMFs to produce 64 QMF subband signals and they are further analyzed by Nyquist filters to increase the resolution of low frequency components. That is, the 3 lowest QMF subbands are split into 10 hybrid subbands by 13-tap finite impulse response (FIR) filters and the other 61 QMF subbands are delay-compensated to compose 61 hybrid subbands. The spatial parameters (CLDs, ICC, and CPCs) are applied in the hybrid subband domain, where the number of hybrid subbands is 71, and the signals are processed by Nyquist and QMF synthesis to produce output multi-channel signals. The time resolution of hybrid subband signals is identical to that of QMF subband signals.

Since the downmix signal and spatial parameters are used together in MPEG Surround decoding, their time synchronization is crucial. The QMF analysis and synthesis processes are an FIR filtering process and produce time delays, which means that the downmix signal in the time-domain connection is delayed compared to that in the QMF subband-domain connection. To simplify the whole decoding structure and minimize an algorithmic delay, MPEG Surround specifies the QMF subband-domain connection as the default mode when it is used with HE-AAC [1]. Once MPEG Surround bit streams are composed by an MPEG Surround encoder so that the downmix signal and spatial parameters are synchronized for the QMF subband-domain connection, the delay compensation process should be applied to the time-domain connection. The process can be performed either by delaying the spatial parameters in the hybrid subband domain or by decoding the downmix signal in advance in the time domain, which we refer to as the spatial parameter-based method (SPBM) and the downmix-based method (DBM) respectively. Since a coding standard should define its decoding process explicitly, the MPEG Surround standard specifies only the SPBM for the time-domain connection [1]. In this paper, we deal with both the SPBM and DBM since they are both implementable for practical applications and their results are identical in the subjective sound quality sense.

In Fig. 1(a), the QMF subband-domain connection is represented as dotted lines. In HE-AAC decoding, the look-ahead of 6 QMF subband samples is available for the 3 low frequency (LF) subbands [5]. Therefore, we extract the LF subband signals including the look-ahead before QMF synthesis in HE-AAC and insert them after QMF analysis in MPEG Surround. In this case, there is actually no additional delay by the following Nyquist analysis in

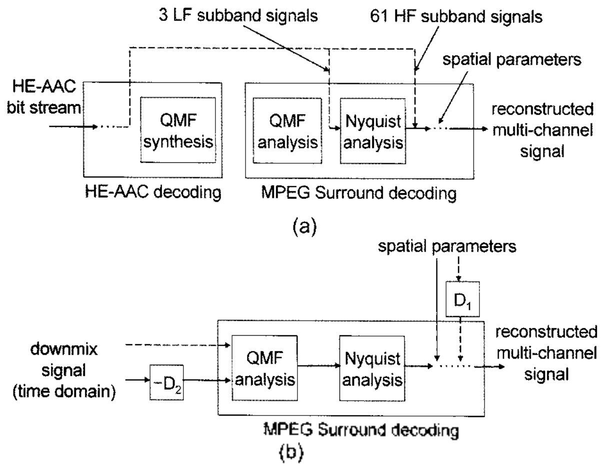

Fig. 1. The dual-domain connection scheme: (a) QMF subband -domain connection method and (b) two time- zdomain connection methods.

MPEG Surround, which is a 13-tap FIR filtering process, because of the look-ahead. Then we extract the 61 high frequency (HF) subband signals before QMF synthesis in HE-AAC and insert them after Nyquist analysis in MPEG Surround. In this case, we can also eliminate the unnecessary delay process required for the HF subband signals in the Nyquist analysis of MPEG Surround. Based on the QMF subband-domain connection, an MPEG Surround encoder synchronizes the spatial parameters with the downmix signal so that there should be no unnecessary delay process in MPEG Surround decoding.

Since the synchronization between the downmix signal and spatial parameters in MPEG Surround bit streams is based on the QMF subband-domain connection, it is corrupted due to an additional delay of the downmix signal for the time-domain connection. The delay is 961 time samples, which is the sum of 257, 320, and 384 time samples caused by QMF synthesis in HE-AAC, QMF and Nyquist analysis in MPEG Surround respectively [1, 5]. As described before, two solutions (SPBM and DBM) are available and shown in Fig. 1(b). In the SPBM, the spatial parameters are delayed by 15 subband samples in the hybrid subband domain, which is represented as dotted lines with D_1=15 subband samples. In the DBM, we decode the downmix signal in advance so that it synchronizes with the spatial parameters in the hybrid subband domain, which is represented as solid lines with D_2=961 time samples.

The SPBM results in a time-misalignment between the downmix signal and spatial parameters since 15 subband samples correspond to 960 (=64*15) time samples, not 961 time samples. The DBM results in an algorithmic delay since it should decode the downmix signal in advance. Therefore, the SPBM can be used for real-time applications where the minimization of an algorithmic delay is important. The DBM can be used for non-real-time applications where the bit-exactness between the output multi-channel signals of the time-domain and QMF subband-domain connections is important.

III. Experimental Results

Experiments have been performed to compare the SPBM and DBM to the QMF subband-domain connection method. We used 11 5.1-channel test items used for MPEG Surround standardization [10], whose lengths were about 20 seconds with a sampling frequency of 44100 Hz. We first encoded the items using an MPEG Surround reference encoder [11] based on 5-1-5 and 5-2-5 configurations in high quality mode, where the bitrates of the MPEG Surround data were about 12 kbps for both the configurations. Note that an M-N-M configuration implies that the numbers of input/output and downmix channels are M and N respectively. The mono or stereo downmix signals were encoded in HE-AAC format with a bitrate of 32 kbps per channel and were further synchronized with the spatial parameters in MPEG Surround bit streams for the QMF subband-domain connection. The frame size of MPEG Surround was 2048 time samples and the number of parameter bands was 28. The downmix signals were decoded to QMF subband and time samples and further used for MPEG Surround decoding. HE-AAC and MPEG Surround were implemented in floating-point source codes and the PCM samples between the codecs were in 16-bit format.

We first performed an informal listening test to compare the output 5.1-channel sounds of the QMF subband-domain connection with those of the SPBM and DBM. As might be expected, they were not subjectively distinguishable at all. Therefore we further analyzed the output 5.1-channel signals of each method from the viewpoint of errors and objective difference grade (ODG) values [12].

We defined the error as differences between the output signals of the QMF subband-domain and time-domain connection methods. Since all the methods have different delay values from each other, they were time-aligned for error calculation. In Table 1, the root mean square errors (RMSEs) and the signal-to-noise ratios (SNRs) are listed, where the noise represents the error between either of the DBM and the SPBM and the QMF subband-domain connection method. The results were first calculated for each channel and then averaged over all the channels. The low frequency enhancement (LFE) channel was excluded since only one item contained non-zero LFE channel signals. The average errors of the DBM are relatively small, where the errors can be explained by two reasons: the impulse response of the QMF analysis and synthesis processes is not a perfect impulse and the output PCM samples of the HE-AAC decoder are quantized in 16-bit format. The maximum error of the DBM was 1~2% of the output of the QMF subband-domain connection method. The errors of the SPBM can be explained by three reasons: the time-misalignment between the downmix signal and spatial parameters by one time sample and the two reasons described in the DBM. As might be expected, the average errors of the SPBM are relatively large mainly due to the time-misalignment. The maximum error of the SPBM was up to 50% of the output of the QMF subband-domain connection method, where the reason of the large error will be explained at the end of the next paragraph.

Table 1. Average RMSEs and SNRs of the SPBM and DBM compared to the QMF subband-domain connection method. The RMSEs represent results for PCM samples in 16-bit format.

Results	DBM (5-1-5)	DBM (5-2-5)	SPBM (5-1-5)	SPBM (5-2-5)
RMSE	7.60	3.38	48.36	37.80
SNR	41.9 dB	47.8 dB	25.8 dB	26.8 dB

For further analysis of the results of the SPBM, we show the error signals calculated for the front-left channels of item 9 in Fig. 2. Though the errors of the SPBM are large, they are highly correlated with the original signal. Assuming that the CLDs have a level of K and are transmitted once per frame and the signal level is L, the error level is approximately $KL/2048$ considering the one-time-sample misalignment. Note that K varies from −45 dB to 45 dB except for extreme cases where K is either −150 dB or 150 dB [1], and L varies from −32768 to 32767. It is noticeable that the errors of the SPBM are relatively large between 6s and 12s, where the original signal has strong high frequency components. Since the variation of high frequency components is large even in the one-time-sample interval, the error increases as the high frequency components become dominant in a signal.

To show that the large errors correlated with the original signals in the SPBM have little effect on the perceived sound quality, we show the ODG values between the QMF subband-domain and time-domain

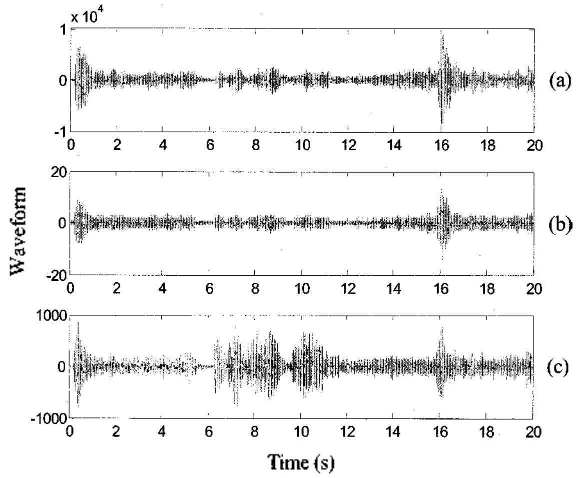

Fig. 2. The front-left channel signals for item 9: (a) output signal of the QMF subband-domain connection method, (b) error signal of the DBM, and (c) error signal of the SPBM.

Table 2. Averaged ODG values with 95% confidence intervals.

ODG values	DBM (5-1-5)	DBM (5-2-5)	SPBM (5-1-5)	SPBM (5-2-5)
upper bound	0.052	0.052	0.005	0.022
mean	0.05	0.05	−0.03	0
lower bound	0.048	0.048	−0.065	−0.022

connection methods in Table 2.

The ODG values were calculated using the advanced mode of the perceptual evaluation of audio quality (PEAQ) technique [12]. In fact, the calculation procedure of the ODG values is not defined for 5.1-channel signals. On the assumption that the subjective sound quality of each channel signal of the SPBM and DBM is little degraded compared to that of the QMF subband-domain connection method, we first calculated the ODG values on a channel basis and then averaged them over all the channels. ODG values range from 0.07 to −4, where a negative value represents sound quality degradation in the mean opinion score (MOS) [12]. Note that a positive value corresponds to a case that a subject cannot discriminate the reference item from the test item [12]. The results are not significantly less than 0 at a 95% confidence level with their mean values very close to 0, which means that the DBM and SPBM do not degrade the perceived sound quality compared to the QMF subband-domain connection method.

IV. Conclusion

We presented the detailed technical description and experimental results of the dual-domain connection scheme for HE-AAC and MPEG Surround, which has not been much investigated yet. Experimental results verify that the SPBM, which is specified as the default time-domain connection method by the MPEG Surround standard, does not cause subjective sound quality degradation compared to the QMF subband-domain connection method. The results also show that the SPBM can be replaced by the DBM especially for non-real-time applications where an algorithmic delay is less important.

Acknowledgment

This work was supported by the faculty research fund of Sejong University in 2008.

References

1. ISO/IEC 23003-1, *Information technology - MPEG audio technologies - Part 1: MPEG Surround*, 2007.
2. J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Linzmeier, and K.S. Chong, "MPEG Surround-The ISO/MPEG standard for efficient and compatible multichannel audio coding," *J. Audio Eng. Soc.*, vol. 56, no. 11, pp. 932-955, 2008.
3. J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen, and S. van de Par, "Background, concept, and architecture for the recent MPEG Surround standard on multi-channel audio compression," *J. Audio Eng. Soc.*, vol. 55, no. 5, pp. 331-351, 2007.
4. G. Hotho, L.F. Villemoes, and J. Breebaart, "A backward-compatible multichannel audio codec," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 83-93, 2008.
5. ISO/IEC 14496-3, *Information technology - Coding of audio-visual objects - Part 3*: Audio, 2005.
6. A. Ehret, M. Dietz, and K. Kjörling, "State-of-the-art audio coding for broadcasting and mobile applications," *AES 114th* convention, paper no. 5834, Mar. 2003.
7. M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," *AES 112nd convention*, paper no. 5553, May 2002.
8. TTAS.KO-07.0027, *Standard for satellite digital multimedia broadcasting transmitter/receiver interface*, 2004.
9. Dong-Soo Kim, Hee-Suk Pang, Jaehyun Lim, Hyen-O Oh, and Yang-Won Jung, "Refined downmix signaling for MPEG Surround," *ISO/IEC JTC1/SC29/WG11*, M13356, 2006.
10. ISO/IEC JTC1/SC29/WG11, *Call for Proposals on Spatial Audio Coding*, N6455, 2004.
11. ISO/IEC JTC1/SC29/WG11, *ISO/IEC 23003-1:2007/FDAM 2, MPEG Surround Reference Software*, N9635, 2008.
12. ITU-R Rec. BS. 1387-1, *Method for objective measurements of perceived audio quality*, 2001.

[Profile]

• Hee-Suk Pang

Hee-Suk Pang received the B.S., M.S., and Ph.D. degrees all from Seoul National University in 1994, 1996, and 2001 respectively. From March 2001 to Feb. 2008, he was a chief research engineer at DM research lab., LG Electronics. Since March 2008, he has been an assistant professor at the department of electronics engineering, Sejong University. His research areas are audio coding, audio signal processing, and music signal analysis.