Clipping Prevention Scheme for MPEG Surround

Hee-Suk Pang

ABSTRACT—MPEG Surround has a potential clipping problem since its encoding is based on downmixing a multichannel signal. We propose a clipping prevention scheme for MPEG Surround, which is composed of modification and recovery processes of a downmix signal with recovery information conveyed in arbitrary downmix gains of an MPEG Surround bitstream. Experiments show that the proposed scheme effectively prevents sound quality degradation caused by clipping problems with negligible additional bit rates.

Keywords—Clipping prevention, MPEG Surround, arbitrary downmix gain (ADG).

I. Introduction

A simple multichannel audio coding method is to compress each channel signal [1]. It guarantees good sound quality but requires a high bit rate, such as 64 kbps per channel. Another multichannel audio coding method based on matrix-based downmix and upmix processes [2] requires a low bit rate, but its sound quality is relatively poor due to information losses in the downmix process.

To achieve good sound quality and a low bit rate, MPEG Surround uses a downmix signal and spatial parameters (SPs), such as channel level differences (CLDs), inter-channel correlation (ICC), and channel prediction coefficients (CPCs) to reconstruct a multichannel signal [3]. Since its bit rate for SPs is about 12 kbps in a normal mode, MPEG Surround is especially effective for broadcasting services such as multichannel digital audio broadcasting. A mono or stereo downmix signal is also available if SPs are ignored.

Since MPEG Surround encoding is based on downmixing a multichannel signal, it has a potential clipping problem that causes sound quality degradation, perceived as a timbre change or a noise addition. To prevent this, a scheme was proposed which modifies the standard MPEG Surround bitstream [4]. We propose a clipping prevention scheme based on the standard MPEG Surround bitstream, which is practically useful for realtime broadcasting and streaming services. We focus on 5-2-5 and 5-1-5 configurations in MPEG Surround, where input and output signals have 5.1 channels, and a downmix signal is either stereo or mono.

Manuscript received Mar. 24, 2008; revised May 6, 2008; accepted June 16, 2008.

Hee-Suk Pang (phone: + 82 2 3408 3727, email: hspang@sejong.ac.kr) is with the Department of Electronics Engineering, Sejong University, Seoul, Rep. of Korea.

II. Clipping Problem in MPEG Surround

In MPEG Surround encoding, a multichannel signal is downmixed into a mono or stereo signal. For a 5.1-channel signal, a mono downmix signal is calculated as

$$x_M = \frac{1}{a_d}\left(x_{FL} + x_{FR} + x_C + \frac{1}{a_s}x_{SL} + \frac{1}{a_s}x_{SR} + \frac{1}{a_{lfe}}x_{LFE}\right), \quad (1)$$

and a stereo downmix signal is calculated as

$$\begin{aligned} x_L &= \frac{1}{a_d}\left(x_{FL} + \frac{1}{\sqrt{2}}x_C + \frac{1}{a_s}x_{SL} + \frac{1}{\sqrt{2}a_{lfe}}x_{LFE}\right), \\ x_R &= \frac{1}{a_d}\left(x_{FR} + \frac{1}{\sqrt{2}}x_C + \frac{1}{a_s}x_{SR} + \frac{1}{\sqrt{2}a_{lfe}}x_{LFE}\right), \end{aligned} \quad (2)$$

where *M*, *FL*, *FR*, *C*, *SL*, *SR*, *LFE*, *L*, and *R* represent mono, front-left, front-right, center, surround-left, surround-right, low frequency enhancement, left, and right channels, respectively. Here, a_d, a_s, and a_{lfe} represent downmix, surround, and LFE gains and are set to predefined values which are equal to or greater than 1 [3]. For example, 1, $\sqrt{2}$, and $\sqrt{10}$ are default values of a_d, a_s and a_{lfe} for the MPEG Surround reference encoder [5]. Once these gains are determined, they are fixed throughout the entire samples. After downmix encoding and decoding by a conventional audio codec, the downmix samples are amplified for compensation of the attenuation by these gains in MPEG Surround decoding.

It follows from (1) or (2) that a downmix signal has a potential clipping problem. Figure 1 shows a 16-bit stereo

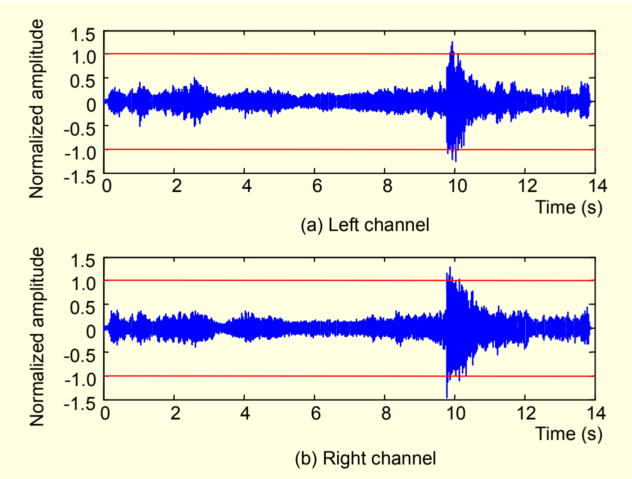

Fig. 1. Example of a clipping problem in a downmix signal.

downmix signal based on (2) with the default setting, where clipping around 10 s degrades the sound quality of both the downmix and reconstructed multichannel signals.

A simple approach to solve this problem is to use a large value of a_d, but this reduces the signal-to-noise ratio (SNR). Especially for real-time broadcasting applications, the SNR must be excessively reduced since the optimal value of a_d cannot be estimated in advance. Another approach is to use a limiter for a downmix signal, but this unnecessarily reduces the dynamic range of a reconstructed multichannel signal compared to that of an original multichannel signal.

III. The Proposed Scheme

The proposed scheme modifies a downmix signal before downmix encoding and recovers it after downmix decoding with recovery information conveyed in arbitrary downmix gains (ADGs), which are an optional tool of MPEG Surround originally used to modify the gains of subband samples [3]. The scheme conserves the dynamic range of an original multichannel signal in the reconstructed multichannel signal.

The basic idea of the proposed encoding scheme is shown in Fig. 2. Suppose the number of ADG sets is L ($L \geq 1$) and ADG^j_m represents the j-th ($j \leq L$) ADG set for the m-th frame. We define a clipping gain, CG_m, as the minimum ADG value required to prevent clipping for the m-th frame. In the proposed scheme, the time positions where SPs are applied are given and uncontrollable. This is because the positions are mainly determined considering CLDs, ICC, and CPCs rather than ADGs and are then shared by all parameters. To calculate ADG^j_m, we first calculate CG_m and CG_{m+1}. When $CG_m < CG_{m+1}$, the last sample of the m-th frame should be attenuated by CG_{m+1} to prevent clipping in the (m+1)th frame, which results in $ADG^j_m = CG_m$ for $j<L$ and $ADG^L_m = CG_{m+1}$. If $CG_m \geq CG_{m+1}$,

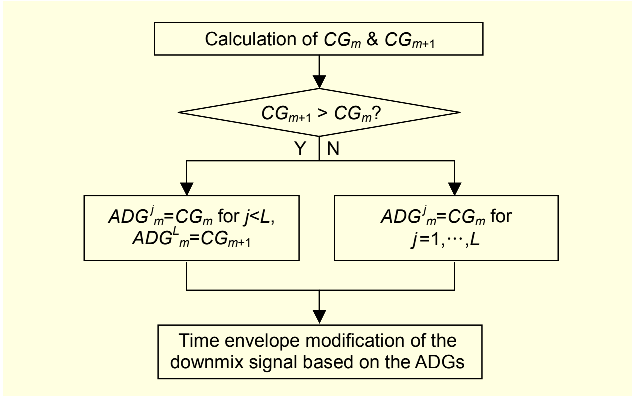

Fig. 2. Proposed encoding scheme.

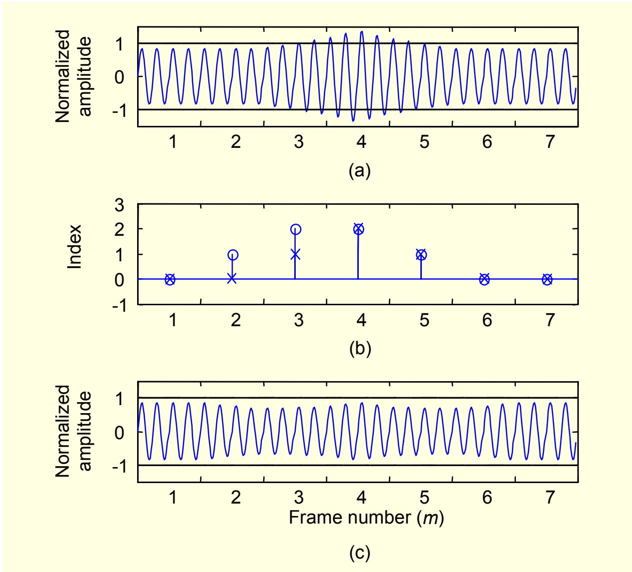

Fig. 3. Example of a downmix signal and the relevant ADGs in the subband domain: (a) original downmix signal, (b) CGI_m (×) and $ADGI^1_m$ (○), and (c) modified downmix signal.

the last sample of the m-th frame is already sufficiently attenuated by CG_m; thus, $ADG^j_m = CG_m$ for $j \leq L$. Then, the time envelope of the downmix signal is linearly attenuated as a function of time in the hybrid subband domain based on the ADGs. The encoding scheme requires one look-ahead frame.

In the decoding process, the ADGs are read from a bitstream, and the time envelope of the downmix signal is recovered based on the ADGs in the hybrid subband domain. Since this process automatically works in a normative MPEG Surround decoder, no modification is required on the decoder side. Further MPEG Surround decoding is performed using the recovered downmix signal.

An example using the proposed encoding scheme is shown in Fig. 3. The number of parameter sets was set to 1 (L=1) for all the frames and the ADGs were applied to the final subband sample in each frame, which is the most frequently used position by the MPEG Surround reference encoder. Since the ADG indexes are conveyed in the bitstream, we define the indexes of CG_m and ADG^j_m as CGI_m and $ADGI^j_m$, respectively. The downmix signal in Fig. 3(a) leads to CGI_3=1, CGI_4=2, CGI_5=1, and CGI_m=0 elsewhere. Then, $ADGI^1_2$=1, $ADGI^1_3$= $ADGI^1_4$=2, $ADGI^1_5$=1, and $ADGI^1_m$=0 elsewhere. The modified downmix signal has no clipping as seen in Fig. 3(c).

Since the proposed scheme focuses on conserving the dynamic range of a reconstructed multichannel signal, a downmix signal is inevitably modified. This modification is primarily based on dynamic range control, which has been used for conventional audio [6].

IV. Experimental Results

From a database, we selected five 5.1-channel test items which show obvious sound quality degradation due to clipping problems. The signals are from 15 to 20 s in duration with a sampling frequency of 44,100 Hz. We used 5-1-5 configurations for items 1 to 4 and a 5-2-5 configuration for item 5, where a_d, a_s, and a_{lfe} were set to the default setting. MPEG4 advanced audio coding (AAC) was used as a downmix codec at a bit rate of 80 kbps per channel. The bit rates of MPEG Surround data were about 12 kbps.

A listening test was performed following the MUSHRA procedure [7], in which each test set is composed of 5 signals. The signals include an anchor, a hidden reference, and the output signals of MPEG Surround with and without the proposed scheme. Since the proposed scheme is based on the ADGs, its minimum gain control step is 2 dB [3]. For comparison, we added a comparative scheme which is generally identical to the proposed scheme except that it uses a gain control step of 0.75 dB and 8 look-ahead frames to prevent an abrupt gain change.

For a test setup, 5.1-channel active loudspeakers were used. Twenty-one listeners with normal hearing participated in the listening test, and the results obtained after the post-screening procedure from 19 listeners are shown in Fig. 4. The proposed scheme is significantly better than normal MPEG Surround without the proposed scheme, at a 95% confidence level. There is no significant difference between the proposed and comparative schemes. For the proposed scheme, the bit rate increases by 49.6 bps, 55.3 bps, 47.4 bps, 119.8 bps, and 144.2 bps for items 1 to 5, respectively. These values are less than or around 1% of the bit rates of MPEG Surround data.

Although the proposed scheme focuses on the sound quality of reconstructed 5.1-channel signals, we also conducted an informal listening test to assess the sound quality of downmix signals. Whereas normal MPEG Surround produced downmix signals with uncomfortable noises due to clipping problems, the proposed scheme did not.

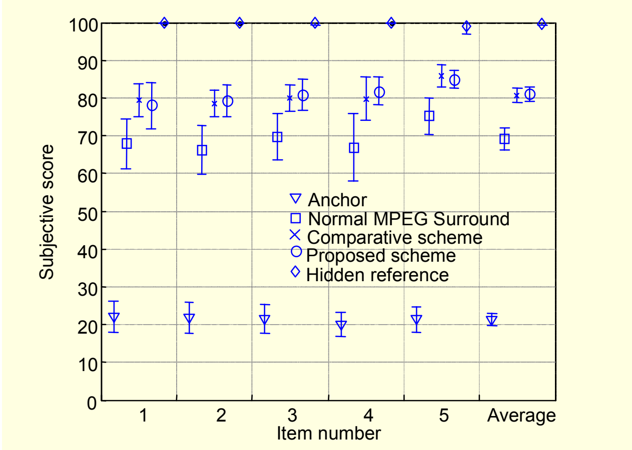

Fig. 4. Result of the listening test at a 95% confidence level.

V. Conclusion

We proposed a clipping prevention scheme for MPEG Surround which is very effective and leads to negligible increases in bit rate. Since the proposed scheme is based on the standard MPEG Surround bitstream, it is practically useful.

References

[1] M. Bosi, "High-Quality Multichannel Audio Coding: Trends and Challenges," *J. Audio Eng. Soc.*, vol. 48, no. 6, 2000, pp. 588-595.

[2] M.F. Davis and M.C. Fellers, "Virtual Surround Presentation of Dolby AC-3 and Pro Logic Signals," *AES 103rd Convention*, Preprint no. 4542, 1997.

[3] ISO/IEC 23003-1, *Information Technology–MPEG Audio Technologies – Part 1: MPEG Surround*, 2007.

[4] H.S. Pang et al., "Clipping Prevention Scheme for MPEG Surround," *ISO/IEC JTC1/SC29/WG11*, M13355, Apr. 2006.

[5] ISO/IEC JTC1/SC29/WG11, *ISO/IEC 23003-:2007/FPDAM 2, MPEG Surround Reference Software*, N9093, 2007.

[6] W. Hoeg et al., "Dynamic Range Control (DRC) and Music/ Speech Control (MSC)–Programme–Associated Data Services for DAB," *EBU Technical Review*, no. 261, 1994, pp. 56-70.

[7] ITU-R BS 1534-1, *Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems*, ITU, 2003.