Improved Channel Level Difference Quantization for Spatial Audio Coding

Kwangki Kim, Seungkwon Beack, Jeongil Seo, Daeyoung Jang, and Minsoo Hahn

ABSTRACT—The channel level difference (CLD) is a main parameter in the reference model 0 (RM0) for MPEG Surround. Nevertheless, the CLD quantization method in the RM0 has problems such as the lack of theoretical background and inappropriate quantization levels. In this letter, a new CLD quantization method is proposed based on the virtual source location information which has strength in the quantization process. From experimental results, it is confirmed that the proposed scheme greatly reduces the quantization distortions measured in dB and degrees without any additional complexity.

Keywords—MPEG Surround, channel level difference (CLD), virtual source location information (VSLI), multi-channel audio.

I. Introduction

Recently, the MPEG Audio sub-group has been working for the standardization of MPEG Surround which is a technology to encode and decode multi-channel audio signals effectively [1], [2]. The scheme of MPEG Surround is similar to that of binaural cue coding (BCC) to compress the multi-channel audio signals as mono or stereo down-mixed signal and spatial cues [3].

Reference model 0 (RM0) for MPEG Surround mainly uses channel level difference (CLD) and inter-channel correlation (ICC) as spatial cues. The CLD and the ICC are defined as the level difference and the amount of correlation or coherence between two channels, respectively. The CLD is a main parameter in the RM0 for MPEG Surround because it determines the spectral power of the reconstructed multi-channel audio signals and occupies a considerable amount of the side information. For this reason, CLD quantization must be reliable to maintain original sound source quality.

The CLD quantization method in the RM0 is a heuristically extended version of the interaural intensity difference (IID) quantization method in parametric stereo coders and has inappropriate quantization levels because of its wide dynamic range. Consequently, the CLD quantization method could result in quantization errors and sound quality degradation. In this letter, we present a new CLD quantization method based on virtual source location information (VSLI) [4]-[6]. It not only has a theoretically more reliable background but also robustness to quantization distortion.

Manuscript received Aug. 09, 2006; revised Sept. 20, 2006.

Kwangki Kim (phone: +82 42 866 6204, email: kkkim@icu.ac.kr) and Minsoo Hahn (email: mshahn@icu.ac.kr) are with the School of Engineering, Information and Communications University, Daejeon, Korea.

Seungkwon Beack (email: skbeack@etri.re.kr), Jeongil Seo (email: seoji@etri.re.kr), and Daeyoung Jang (email: dyjang@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

II. Problem Description

The CLD is generally estimated by the power ratio between the reference and the other compared channel signal in each frequency and time bin. Then, this ratio is converted to a logarithm scale. Since the CLD is a main parameter which is used to reconstruct the power gain of each output channel, it is indispensable to quantize the CLD to transmit the CLD information. Due to the CLD quantization with a limited number of bits, CLD distortion is unavoidable and results in the degradation of the reconstructed audio sound quality.

In order to quantize the CLD, the RM0 exploits a non-uniform bidirectional 31 level quantizer with a limited dynamic range of $\pm$150 dB. In this CLD quantization method, small CLD values tend to be quantized precisely while large CLD values are quantized rather roughly.

Figure 1 shows the CLD distribution of 11 multi-channel audio samples which were offered by the MPEG audio sub-group [2]. In this figure, it can be confirmed that most CLD values exist within a range of 30 dB and CLD distribution beyond this 30 dB range is negligible. Therefore, the dynamic range of the CLD quantization method should be reduced and small CLD values should be quantized more finely. However, due to unreasonable quantization levels, RM0 suffers from CLD quantization errors. Moreover, the current CLD quantization method is not able to preserve the general human perceptual criterion. In other words, a less than 3 dB level difference resolution must be guaranteed to maintain a perceptually permissible spectral distortion [7].

A CLD quantization method using the spatial image of multi-channel sounds is a possible solution. Actually, because the key role of spatial cues is to keep the spatial image of the sounds, the VSLI which is the angle information of the spatial image and has the strength in the quantization process can be directly used to quantize the CLD rather successfully. The details are described in the next chapter.

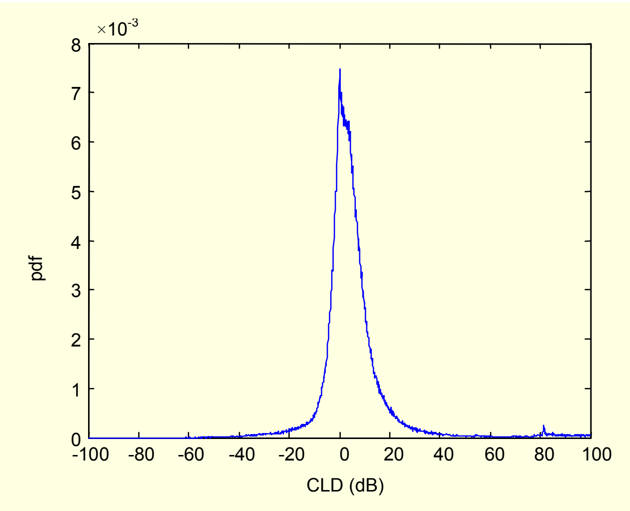

Fig. 1. CLD distribution of 11 multi-channel audio samples.

III. Proposed CLD Quantization Method

The VSLI originates from the fact that the spatial image between a pair of channels can be represented as a panning angle and the power amplitude using the constant power panning law. The information of the pair of channels can be reconstructed from the panning angle and the power amplitude [8]. This concept can be specifically explained in the case of stereophonic signals. It is assumed that stereophonic signals are virtually localized in the edge of a sector whose angle is 90 degrees. The spatial image of two channel signals can be represented by the VSLI to be the angle information and the downmixed signal to be the power amplitude information. The VSLI in partition b, denoted by $VSLI_b$ is estimated as

$$VSLI_b = \tan^{-1}\frac{\sqrt{P_{r,b}}}{\sqrt{P_{l,b}}}, \qquad (1)$$

where $P_{l,b}$ and $P_{r,b}$ are the estimated power of the left and right signals, respectively, while b is a parameter band index which indicates the spectral region containing several spectral components [3]. By using the estimated $VSLI_b$ and the power amplitude of the downmixed signal, $P_{d,b}$, the stereophonic signals are inversely reconstructed as

$$\begin{aligned} P_{l,b} &= P_{d,b} \times (\cos(VSLI_b))^2, \\ P_{r,b} &= P_{d,b} \times (\sin(VSLI_b))^2. \end{aligned} \qquad (2)$$

Using (2) and the definition of the CLD, the relationship between the VSLI and the CLD can be derived as

$$CLD_b = 10\log_{10}\frac{P_{r,b}}{P_{l,b}} = 10\log_{10}(\tan(VSLI_b))^2, \qquad (3)$$

where CLD_b is the CLD in partition b.

Meanwhile, the VSLI should be also quantized and encoded for transmission. Since it is well known that a human being cannot easily distinguish variation of the spatial image of an audio signal within 3 degrees of distortion [7], a 31 level uniform quantizer which has a three-degree resolution can be used for the VSLI quantization with the dynamic range of 90 degrees. Thus, the VSLI can be quantized without any perceptual degradation. This strength makes the VSLI superior to the CLD with respect to quantization.

Based on the strength of the VSLI quantization and the relationship between the VSLI and the CLD, the CLD quantization method in the RM0 can be replaced by the VSLI-based CLD one. In other words, the quantization levels and the decision levels of the VSLI quantization scheme can be directly converted into those of the CLD one using (3). Figure 2 illustrates the schematic derivation of the CLD quantization levels and the decision levels from the VSLI. As shown in Fig. 2, the proposed CLD quantization method has both non-uniform quantization levels and non-uniform decision levels. Figure 3 shows the CLD quantization levels of the original and the proposed methods in dB and degrees.

The proposed quantization method has fine resolution in the region of small CLD values as shown in Fig. 3(a) and focuses on the effective representation of the spatial image as indicated by the data shown in Fig. 3(b). In degrees, while the original CLD quantization method has non-uniform quantization and decision levels, the proposed method has uniform quantization and decision levels. Therefore, the proposed CLD quantization method is more reasonable than the original method. Finally, by using the proposed method, it is expected that the CLD

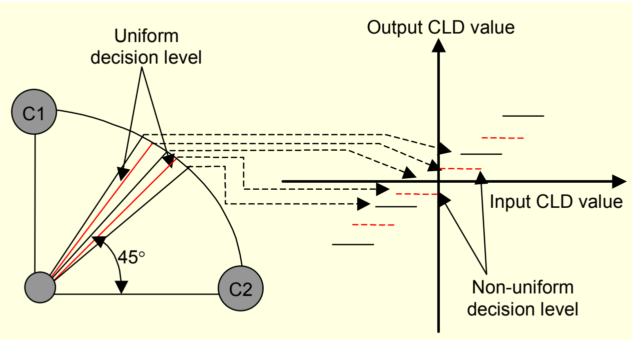

Fig. 2. Schematic derivation of the CLD quantization and decision levels from the VSLI.

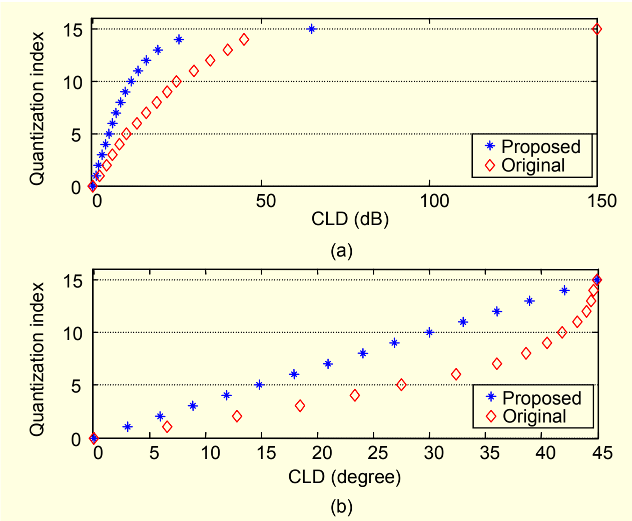

Fig. 3. Comparison of the CLD quantization levels: (a) in dB, (b) in degrees.

quantization distortions in dB and degrees can be reduced.

IV. Experimental Results

To confirm the effectiveness of our new method, comparison tests were conducted on the aspect of quantization error in dB and degrees as quantitative measures. Subjective audio quality assessment was also performed. For all tests, the eleven test sequences offered by MPEG Audio sub-group were used [2]. The items were sampled at 44.1 kHz with 16 bit resolution and were all shorter than 20 seconds. These test items are listed in Table 1.

Figure 4 shows a comparison between the quantization errors of the original and the proposed method for all test items with different CLD values. The total average quantization errors are shown in Table 2. From Fig. 4, it is clear that the proposed quantization method has smaller quantization errors both in dB and degrees. Moreover, as shown in Table 2, the average quantization errors of the proposed method are about 50% smaller than those of the original method. However, the proposed quantization method has a slightly higher bit rate than the original method. This is because MPEG Surround adopts Huffman coding and the proposed method has fine resolutions which result in an increase of bit rate in the Huffman coding.

As a subjective listening test, a MUSHRA test was

Table 1. Performance evaluation materials.

Index	Material	Description
A	Applause	Ambience
B	ARL_applause	Ambience
C	Chostakovitch	Music (back: direct)
D	Fountain_music	Pathological
E	Glock	Pathological
F	Indie2	Movie sound
G	Jackson1	Music (back: ambience)
H	Pops	Music (back: direct)
I	Poulenc	Music (back: direct)
J	Rock_concert	Music (back: ambience)
K	Stomp	Movie sound

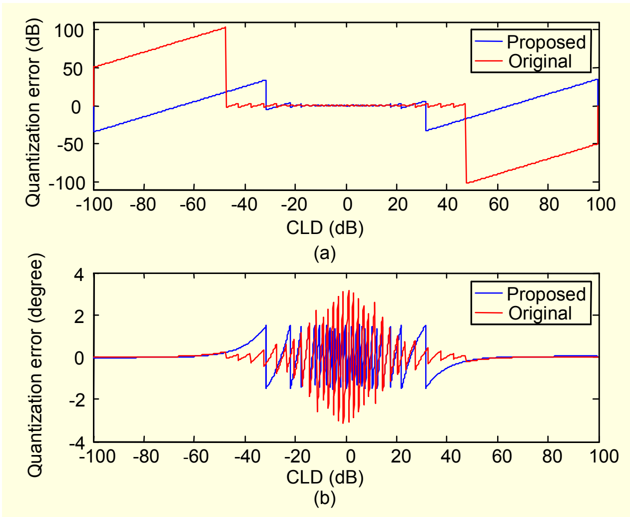

Fig. 4. CLD quantization errors according to the CLD values: (a) in dB, (b) in degrees.

Table 2. Average quantization errors.

Method	Distortion (dB)	Reduction rate (%)	Distortion (degree)	Reduction rate (%)
Original	10.19	55.05	1.55	44.52
Proposed	4.58		0.86	

Table 3. Five multi-channel audio systems tested.

Classification	Description
Hidden reference	Original multi-channel audio signals
Anchor	3.5 kHz band-limited
Additional anchor	Software-based commercial Dolby ProLogic II
RM0	RM0 with the original CLD quantization
Proposed	RM0 with the proposed CLD quantization

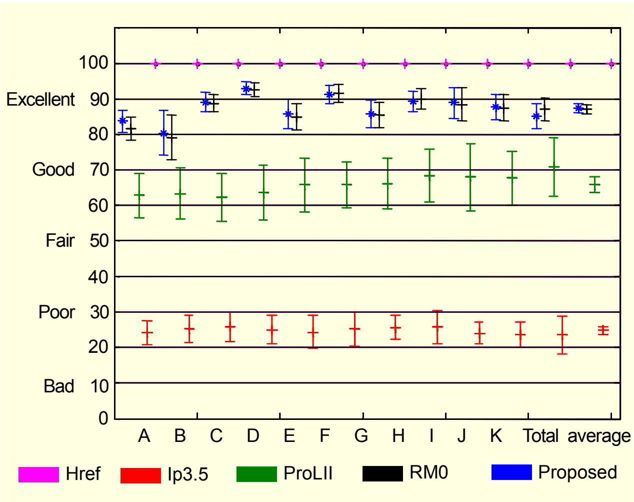

Fig. 5. Subjective listening test results.

performed [9]. Five systems were used for the test and they are listed in Table 3. Ten experienced listeners evaluated the decoded audio quality of the test items in each trial. Figure 5 shows the subjective listening test results. Subjective listening test results show that the sound quality difference between the original and the proposed quantization is negligible. Because the current MPEG Surround coder uses other signal manipulations in order to enhance the sound quality, the effect of the CLD on the sound quality becomes very small. Thus, to confirm the influence of the proposed quantization method on the sound quality, another listening test should be performed using a basic coder which does not use any other additional coding schemes.

V. Conclusion

This letter presents a new CLD quantization method based on the VSLI which has strength in the quantization process and can be directly converted into the CLD. Due to the theoretically motivated background and the appropriate quantization levels, the proposed CLD quantization scheme is more reliable and efficient than the original CLD method. From the experimental results, it is confirmed that the newly proposed quantization scheme greatly reduces the quantization errors both in dB and degrees. In spite of very small quantization errors, the proposed method yields no significant improvement in the sound quality because the current MPEG Surround RM0 coder employs other signal manipulations. To validate the proposed quantization method, another subjective listening test which focuses on the effect of the CLD itself on sound quality should be performed. Additionally, because the proposed quantization scheme is based on the spatial image of multi-channel sound, a listening test which concentrates on the perception of the spatial image also remains as a subject for future work.

References

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), *Call for Proposal on Spatial Audio Coding*, Document N6455, Mar. 2004.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), *Procedures for the Evaluation of Spatial Audio Coding Systems*, Document N6691, Redmond, July 2004.

[3] C. Faller and R. Baumgarte, "Binaural Cue Coding–Part II: Schemes and Application," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

[4] H. Moon, *A Multi-Channel Audio Compression Method with Virtual Source Location Information for MPEG-4 SAC*, Seoul National Univ. Ph.D. Dissertation, Feb. 2005.

[5] S. Beack, J. Seo, H. Moon, K. Kang, and M. Hahn, "Angle-Based Virtual Source Location Representation for Spatial Audio Coding," *ETRI J.*, vol. 28, no. 2, Apr. 2006, pp. 219-222.

[6] J. Seo, H. Moon, S. Beack, K. Kang, and J. 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.

[7] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, Berlin, Heidelberg, 1999.

[8] J.R. West, *Five-Channel Panning Laws: An Analytic and Experimental Comparison*, Master's Thesis, Music Engineering, University of Miami, 1998.

[9] ITU-R Recommendation, *Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)*, ITU, BS. 1534-1, Geneva, 2001.