An Efficient Representation Method for ICLD with Robustness to Spectral Distortion

Seungkwon Beack, Jeongil Seo, Kyungok Kang, and Minsoo Hahn

ABSTRACT—The inter-channel level difference (ICLD) is a cue parameter to estimate spectral information in binaural cue coding that has been recently in the spotlight as a multichannel audio signal compression technique. Even though the ICLD is an essential parameter, it is generally distorted by quantization. In this paper, a new modified ICLD representation method to minimize the quantization distortion is proposed by adopting a flexible determination of the reference channel and the unidirectional quantization scheme. Our experimental result confirms that the proposed method improves the multichannel audio output quality even with the reduced bit-rate.

Keywords—Spatial audio coding, binaural cue coding, multi-channel audio.

I. Introduction

Spatial audio coding (SAC) is a process to represent multichannel audio signals as a down-mixed signal with spatial cues. The main strength of SAC is the significant bit-rate reduction while maintaining the perceptual sound quality. Binaural cue coding (BCC) has been introduced and is now becoming an important scheme for multichannel SAC both in the sense of audio coding and in the standardization issue in MPEG [1]-[6]. In this scheme, spatial cues are defined as augmented (or side) information including the inter-channel level difference (ICLD), the inter-channel time difference (ICTD), and the inter-channel coherence (ICC). These cues play a pivotal role to maximize the similarity between the original and output signals. Therefore, the reconstruction of the original signal from the down-mixed one can be achieved by modifying the spectral power, the phase, and the spatial image diffuseness using the ICLD, the ICTD, and the ICC, respectively.

However, there are some intrinsic limitations. One of the limitations, the main topic of this paper, is that the quantized spatial cues for transmission may lead to a decoded signal degradation. It is natural that more roughly quantized spatial cues result in more distorted outputs. Presently, the required total bit-rate for the augmented information in SAC is about 20 kbps [5]. In this paper, to minimize the output power spectral distortion, an efficient and reliable ICLD quantization method is exploited. Our goal is achieved by some modifications, and it is confirmed that the proposed method is superior in the sense of the lower spectral distortion to the conventional ICLD (C-ICLD) method.

II. Problem Description of Quantized C-ICLD

The ICLD approximation by quantization inevitably causes the spectral distortion of output signals. It is believed that when a less than 3 dB level difference resolution, which is the perceptual level under simultaneous tones [7], is guaranteed, the spectral distortion is perceptually permissible. Another restriction to be considered is that the bit-rate for the augmented information is about 20 kbps when each channel signal is encoded by the conventional AAC coder with 64 kbps.

To satisfy the above considerations, a conventional BCC coder uses a uniform bidirectional 16 level quantizer with the limited dynamic range of ±18 dB [1]. The quantization process seems to satisfy the less than 3 dB resolution condition, but it is successful only when one pair of simultaneous tones are represented by one ICLD between the fixed reference and the other target channels. In this case, the accumulated

Manuscript received Mar. 15, 2005; revised Apr. 16, 2005.

Seungkwon Beack (phone: +82 42 866 6296, email: skbeack@icu.ac.kr) and Minsoo Hahn (email: mshahn@icu.ac.kr) are with Multimedia Group, ICU, Daejeon, Korea.

Jeongil Seo (email: seoji@etri.re.kr) and Kyungok Kang (email: kokang@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

distortion under the occurrence of all simultaneous tones generated from all channels in one subband may exceed 3 dB, and thus the spatial audio image is likely to be degraded. In addition, the restriction of the fixed dynamic range can produce considerably large quantization errors, i.e., the clipping effect, when the estimated power of the fixed reference is considerably lower than that of the target one.

Figure 1 shows the C-ICLD distribution before quantization for several contents. A rather narrow dynamic range, the symmetric distribution, and the almost zero mean, as shown in Fig. 1(a), are appropriate for a conventional quantization scheme. But the distribution of Figs. 1(b) and 1(c) are asymmetrical with considerable large dynamic ranges with non-zero means, so the conventional bidirectional quantizer is not adequate.

The problems of the C-ICLD quantization process can be said as follows. One is a tradeoff problem, i.e., extension of the ICLD dynamic range coverage requires more bits to satisfy a sufficient resolution, thus it easily violates the recommendation of the bit-rate allocated to the augmented information. The other is that the asymmetrically distributed C-ICLD is not suitable for a bidirectional quantizer since several quantization levels may be rarely used to represent the corresponding ICLD values.

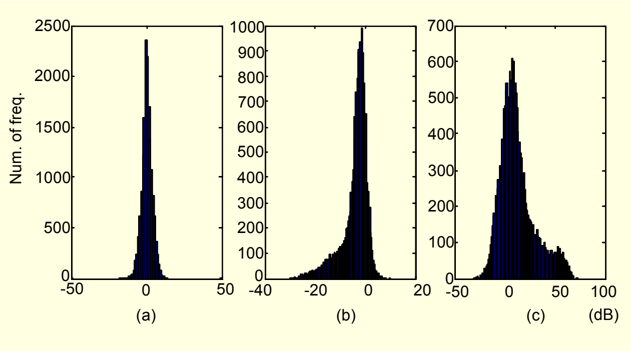

Fig. 1. C-ICLD histograms of five-channel audio contents: (a) applause, (b) classical music, (c) popular music.

III. Proposed Half-Region ICLD (H-ICLD) and Quantization

To reduce the spectral distortion caused by the unsuitably quantized ICLD, a more suitable quantization scheme is proposed. Some modifications using a dominant channel index (DCI) are presented to estimate the ICLD as follows. The index, rf_b, defined as the DCI in partition b, is determined as

$$rf_b = \arg\max_{1 \le c \le C} P_{c,b}, \tag{1}$$

where C is the number of the playback channels and $P_{c,b}$ is an estimated power as

$$P_{c,b} = \sum_{n=A_b}^{A_{b+1}-1} \left| S_{c,n} \right|^2. \tag{2}$$

In (2), $S_{c,n}$ denotes the spectral coefficients of channel c, and A_b is the partition boundary with respect to the equivalent rectangular bandwidth as appeared in [3]. The ICLD can be newly estimated as

$$\Delta L^h_{c,b} = 10\log_{10}\left(\frac{P_{c,b}}{P_{rf_b,b}+\alpha}\right), \quad rf_b > c,$$

$$\Delta L^h_{c-1,b} = 10\log_{10}\left(\frac{P_{c,b}}{P_{rf_b,b}+\alpha}\right), \quad rf_b < c. \tag{3}$$

Here, α is an arbitrary small constant to keep the ICLD stable. We named it an H-ICLD since it always handles only the negative region. In the synthesis process, to convert the H-ICLD into the gain factor $G_{c,b}$, rf_b is used to determine the reference channel in every subband partition. Thus, rf_b information needs to be transmitted. The gain factor $G_{rf_b,b}$ for the partition of the reference channel is

$$G_{ref,b} = \frac{1}{\sqrt{1+\sum_{i=1}^{C-1} 10^{\Delta L^h_{i,b}/10}}}. \tag{4}$$

Using (4), other gain factors can be obtained as

$$G_{c,b} = 10^{\Delta L^h_{c,b}/10} G_{rf_b,b}, \qquad c < rf_b,$$

$$G_{c,b} = 10^{\Delta L^h_{c-1,b}/10} G_{rf_b,b}, \qquad c > rf_b. \tag{5}$$

Each partition gain factor is utilized to reconstruct spectral structures of the corresponding partitions for all channels.

$$U_{c,k} = G_{c,b} S'_k, \quad A_b \le k \le A_{b+1} - 1, \tag{6}$$

where S'_k and $U_{c,k}$ are the spectral coefficients of the down-mixed signal and the output signal of channel c, respectively.

Figure 2 illustrates the H-ICLD distribution before quantization for the same contents as in Fig. 1. Table 1 shows the standard deviation for the C-ICLD and H-ICLD distribution, respectively, according to the distribution in Figs. 1 and 2. Here, two essential characteristics of the H-ICLD, which can be utilized through the quantization process, can be observed: The first characteristic is that the unidirectional

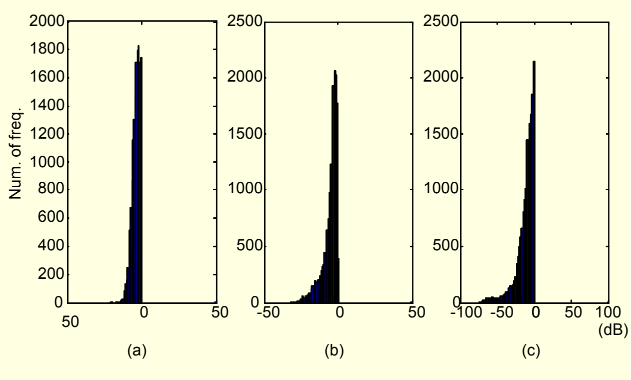

Fig. 2. H-ICLD histograms of five-channel audio contents: (a) applause, (b) classical music, (c) popular music.

Table 1. Standard deviation of C-ICLD and H-ICLD.

Contents Cue (dB)	Ambience sound	Classical music	Popular music
C-ICLD	3.653	5.334	17.148
H-ICLD	2.673	4.806	12.030

distribution decayed from 0 dB can be quantized by a unidirectional quantizer with the half levels of the C-ICLD's quantizer. Second, since the deviation of the H-ICLD is commonly smaller than that of the C-ICLD, the H-ICLD would be more efficiently quantized within the same limited dynamic range. Utilizing these properties, our quantization method is designed. By the unidirectional property, an asymmetrical midtread quantization scheme including the zero level is adopted, so the total number of levels is even. For instance, when Q is 7, quantizer indices are from 0 to 7, and the quantizer indices can be obtained by

$$I_{c,b} = \left\lfloor \frac{-\Delta L^{h}_{c,b}\cdot Q}{\Delta L_{\max}} \right\rfloor, \qquad (7)$$

where $\Delta L_{\max}$ is a limited dynamic range and the H-ICLD is converted into a positive value to obtain the quantizer indices of Q={0,1,…,Q}. The quantizer indices together with the shifted DCI, i.e., shifted by rf_b–1, are encoded with an entropy coder. At the decoder side, after the bitstream of the augmented information is decoded by an entropy decoder, the H-ICLD is obtained using the quantizer indices as

$$\Delta \overline{L}^{h}_{c,b} = -\frac{\Delta L_{\max}\cdot I_{c,b}}{Q}, \qquad (8)$$

and finally 1 is added to the shifted DCI. By utilizing the above characteristics, it can be expected that the quantization distortion and the degradation of the output signal, in the case of simultaneous tones, can be improved.

IV. Experimental Results

Objective tests were conducted to prove that our proposed H-ICLD is more robust for the spectral distortion than the C-ICLD. All the eleven test items offered by the MPEG audio group were used for our tests [6]. All these items have five channels with a 44.1 kHz sampling rate and 16-bit quantization level. Our analysis window for the Fourier transform and the analyzed hop size are identical to [3].

As an objective measurement, we measured the spectral distance between the references and the reconstructed signals. Even though it is not so clear that the degree of the spectral distortion is entirely related to the degree of the perceptual degradation, it would be better to provide a rather objective measure for a more generalized comparison. The symmetric Kullback-Leibler distance is a well known spectral distance measure and its distance is especially highly correlated to the perceptual degradation, while it also estimates the audible spectral discontinuities rather reliably [8]. The symmetric Kullback-Leibler distance is defined as,

$$D_{SKL} = \int (P(\omega) - Q(\omega)) \log \frac{P(\omega)}{Q(\omega)} d\omega. \qquad (9)$$

Here, $P(\omega)$ and $Q(\omega)$ are the power spectra of the reference and decoded signal, respectively. Several accumulative D_{SKL} values for all items are calculated from the reconstructed signals, which are decoded by using the quantized C-ICLD, the quantized H-ICLD, and the unquantized ICLD whose accumulated D_{SKL} is denoted as a minimal distortion boundary. The quantization level is varied from 7Q to 31Q in our test. In the case of the C-ICLD, the quantization scheme in [1] is adopted. Since more than 31Q does not guarantee low bit-rates, the experimental results only for below 31Q cases are depicted. D_{SKL} is also measured with respect to several dynamic ranges of 18, 24, and 30 dB.

From Fig. 3, the results of our objective test can be summarized as follows: First, the quantized H-ICLD consistently produces lower distortion for the same limited dynamic range than that of the quantized C-ICLD. Surprisingly, the H-ICLD at a 30 dB dynamic range almost approaches the minimal distortion boundary. Second, the distortion with the H-ICLD changes very little for various quantization levels. Finally, it can be said that the distortion amount tends to depend heavily on the maximum dynamic range. Unlike [1], it is observed that

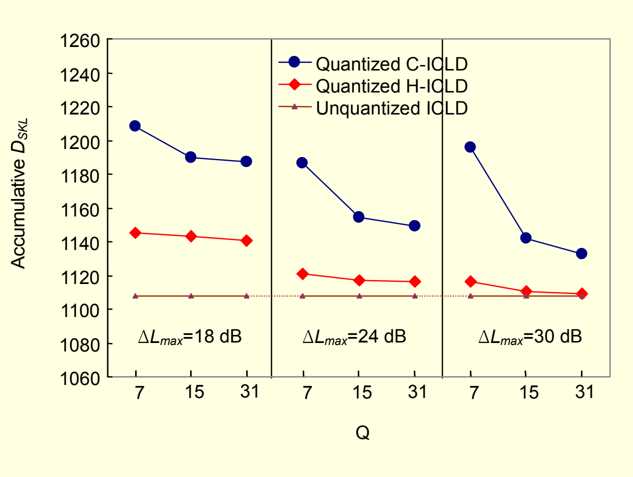

Fig. 3. Accumulative D_{SKL} based performances.

Table 2. Average bit-rate for various quantization levels.

Q(kbps)	7Q	15Q	31Q	63Q
C-ICLD	12.47	18.62	24.86	31.05
H-ICLD	15.77	12.61	27.72	34.05

$\Delta L_{\max}$ should be determined within the range of about 30 dB for both the H-ICLD and the C-ICLD. It should be considered as one of the critical factors to improve the accuracy of the ICLD.

Table 2 shows the comparison of the average bit-rate between the C-ICLD and the H-ICLD for all eleven contents. In order to reduce the bit-rate, the different quantization indices between both consecutive partitions and consecutive frames are encoded by Huffman coding, and the low bit code word is then selected. In order to provide either partition- or frame-based information, one bit is allocated to each codeword.

Due to the auxiliary information of the DCI, the bit-rate of the H-ICLD is slightly increased. Another bit-rate increase comes from the fact that the smoothed sharpness of the H-ICLD within the limited dynamic range requires several more bits. However, careful inspection of Fig. 3 permits us to say that even when the same bit-rate is adopted for both the H-ICLD and the C-ICLD, the output signal-to-noise ratio with the H-ICLD is superior to that with the C-ICLD.

V. Conclusion

The H-ICLD with a new representation method is proposed as a more suitable form to be adopted in the quantization process. It is confirmed that the accuracy of the decoded H-ICLD is better than that of the decoded C-ICLD through the spectral distortion measurements. It can be said that our proposed representation of the H-ICLD is useful to improve the performance of the multichannel audio compression scheme. Namely, it reduces the spectral distortion of the multichannel audio output even at the reduced bit-rates. However, it has to be admitted that an overall subjective test with consideration of all spatial cues including the ICC and the ICTD remains for further study.

References

[1] C. Faller and F. Baumgarte, “Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression,” *AES 112th Convention*, Preprint 5574, May 2002.

[2] C. Faller and F. Baumgarte, “Binaural Cue Coding Applied to Audio Compression with Flexible Rendering,” *AES 113th Convention*, Preprint 5686, Oct. 2002.

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

[4] C. Faller, “Parametric Coding of Spatial Audio,” *Proc. 7th Int. Conf. on Digital Audio Effects*, Naples, Italy, Oct. 2004, pp. 151-156.

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

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

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

[8] E. Klabbers and R. Veldhuis, “Reducing Audible Spectral Discontinuities,” *IEEE Trans. Speech and Audio Proc*. vol. 9, no. 1, Jan. 2001, pp. 39 – 51.