Joint Channel Coding Based on Principal Component Analysis

Dong-il Hyun, Donggeum Lee, Youngcheol Park, Daehee Youn, and Jeongil Seo

This paper proposes a new joint channel coding algorithm based on principal component analysis. A conventional joint channel coder using passive downmixing undergoes a reduction of both the primary-to-ambient energy ratio (PAR) of the downmix signal and the panning gain ratio of the primary source. The proposed system preserves the PAR of the downmix signal by using active downmixing which reflects spatial characteristic. The proposed system also improves the accuracy of the panning gain ratio estimation. Computer simulations and subjective listening tests verify the performance of the proposed system.

Keywords: Spatial audio coding, primary-to-ambient energy ratio, principal component analysis, active downmixing.

I. Introduction

There have been many recent studies [1], [2] on spatial audio coding, the main idea of which is to restore multichannel audio using a downmix signal and spatial parameters which are transmitted by an encoder.

This study demonstrates that a conventional joint channel coder using passive downmixing undergoes a reduction of both the primary-to-ambient energy ratio (PAR) of the downmix signal and the panning gain ratio of the primary source. We propose a new downmixing algorithm based on principal component analysis (PCA) and inter-primary level difference (IPLD), which determines the source's direction. To overcome PCA-based downmixing's distortion problem, we use a weighted IPLD.

II. Conventional Joint Channel Coding

Spatial audio coding systems, such as Parametric Stereo of HE-AACv2 and MPEG Surround (MPS), include single or multiple two-to-one (TTO) modules, which comprise downmixing and spatial parameter estimation.

The two channel signals $X_i[k]$ are often modeled using a primary component $S[k]$ and ambient components $N_i[k]$ which are orthogonal to each other [3]. Then, we can write the input signals as

$$X_1[k] = a_1 S[k] + N_1[k], \quad X_2[k] = a_2 S[k] + N_2[k], \tag{1}$$

where a_1 and a_2 are the respective panning gains of the two channels, which satisfy the relation $a_1^2 + a_2^2 = 1$, and k represents the frequency bin index.

The downmix signal $\tilde{S}[k]$ in the TTO module is generated by simple summation of the input signals as

$$\tilde{S}[k] = (a_1 + a_2) S[k] + N_1[k] + N_2[k]. \tag{2}$$

Thus, the PAR of the downmix signal is calculated as

$$PAR_{\mathrm{MPS}} = \frac{(a_1 + a_2)^2}{2} \cdot \left(\frac{\sigma_S^2}{\sigma_N^2} \right), \tag{3}$$

where σ_S^2 and σ_N^2 are the variances of the primary and ambient components, respectively. Because $(a_1+a_2)^2 \le 2$, this passive downmixing reduces the PAR.

A conventional TTO module estimates two spatial parameters, namely, channel level difference (CLD) and interchannel correlation (ICC), which are calculated by

Manuscript received Apr. 19, 2010; revised June 21, 2010; accepted July 4, 2010.

The work was supported by the IT R&D program (2008-F-011, Development of Next Generation DTV Core Technology) of KEIT&KCC&MKE, Rep. of Korea.

Dong-il Hyun (phone: +82 2 2123 4534, email: fireloaf@dsp.yonsei.ac.kr), Donggeum Lee (email: riglord@dsp.yonsei.ac.kr), Daehee Youn (email: dhyoun@yonsei.ac.kr) are with the Department of Electrical and Electronic Engieering, Yonsei University, Seoul, Rep. of Korea.

Youngcheol Park (email: young00@yonsei.ac.kr) is with the Department of Computer and Telecommunications Engineering, Yonsei University, Wonju, Rep. of Korea.

Jeongil Seo (email: seoji@etri.re.kr) is with the Broadcasting and Telecommunications Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

doi:10.4218/etrij.10.0210.0124

$$CLD[b] = 10\log_{10}\left(R_{11}/R_{22}\right),\ ICC[b] = \left|R_{12}\right|/\sqrt{R_{11}R_{22}}\ . \quad (4)$$

The correlation R_{ij} is calculated as

$$\sum_{m}\sum_{k=k_b}^{k_{b+1}-1} X_i[k]X^*{}_j[k], \quad (5)$$

where b and k_b are the parameter band index and the first frequency bin index of parameter band b, respectively.

The primary level ratio, a_1^2/a_2^2, mainly determines the direction of the phantom source in the input signals. However, because the ambient component affects the CLD parameters of a conventional TTO module, the source direction reconstructed using the CLD can be blurred or even inaccurate. More specifically, the estimated CLD magnitude can be smaller than the panning gain ratio:

$$\left|CLD[b]\right| = \left|10\log_{10}\frac{a_1^2\sigma_S^2+\sigma_N^2}{a_2^2\sigma_S^2+\sigma_N^2}\right| = \left|10\log_{10}\frac{a_1^2 PAR+1}{a_2^2 PAR+1}\right| \le \left|10\log_{10}\frac{a_1^2}{a_2^2}\right|. \quad (6)$$

The conventional ICC estimation is based on a constant phase difference assumption; however, in reality, signals are likely to be rendered with a constant time delay or obtained by binaural recording. Therefore, the constant phase difference assumption is inappropriate for estimating ICC in a band that comprises multiple frequency bins. Furthermore, as the bandwidth of the parameter band widens, the estimated ICCs become increasingly inaccurate [4].

III. Proposed Joint Channel Coding System

1. Active Downmixing Based on PCA

The proposed joint channel coding algorithm uses PCA to separate the primary and ambient components [5]. First, it estimates the correlation matrix of the input channels as

$$\mathrm{R} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}. \quad (7)$$

For practical implementations, sample correlation matrices obtained in frames can be smoothed over several frames [5]. Next, the proposed algorithm calculates the correlation matrix eigenvalues and eigenvectors. Using the signal model in (1), it calculates the input eigenvalues of the correlation matrix as

$$\lambda_0 = \sigma_S{}^2 + \sigma_N{}^2, \quad \lambda_1 = \sigma_N{}^2. \quad (8)$$

It then obtains the downmix signal through a linear combination of the input channels using elements of eigenvector $\mathbf{v_0}$ = $[v_{0,1}\ \ v_{0,2}]^{\mathrm{T}}$, corresponding to the larger eigenvalue λ_0:

$$\tilde{S}[k] = v_{0,1}X_1[k] + v_{0,2}X_2[k] = S[k] + a_1N_1[k] + a_2N_2[k]. \quad (9)$$

From (1), (8), and (9), we can calculate the PAR of the downmix signal straightforwardly as

$$PAR_{\mathrm{PCA}} = \frac{\sigma_S{}^2}{\sigma_N{}^2} = \frac{\left(\lambda_0-\lambda_1\right)}{\lambda_1}. \quad (10)$$

Equation (10) indicates that the active downmixing maintains the PAR regardless of the panning gains a_1 and a_2.

2. IPLD Estimation and Weighing Based on Primary-to-Total Energy Ratio

Using the obtained eigenvalues and eigenvectors, the proposed algorithm estimates IPLD using the following relation:

$$\frac{v_{0,1}}{v_{0,2}} = \frac{R_{12}}{\lambda_0 - R_{11}} = \frac{a_1a_2\sigma_S{}^2}{\left(\sigma_S{}^2+\sigma_N{}^2\right)-\left(a_1{}^2\sigma_S{}^2+\sigma_N{}^2\right)} = \frac{a_1}{a_2}. \quad (11)$$

At the decoder, the downmix signal is repanned using the eigenvector elements. The original source direction, then, is perfectly reconstructed if the downmix signal contains only primary components. However, as (9) shows, the downmix signal contains both primary and ambient components. Thus, both the primary and ambient components are panned at the decoder, which generates unnatural ambience at the decoder. This problem is worse when there are multiple primary components or when there is no clear primary component. To alleviate this problem, we propose a weighted IPLD given by

$$IPLD_{\mathrm{weighted}} = PTR \cdot 20\log 10\frac{v_{0,1}}{v_{0,2}}, \quad (12)$$

where the primary-to-total energy ratio (PTR) is defined as

$$PTR = \frac{\sigma_S{}^2}{\sigma_S{}^2 + 2\sigma_N{}^2} = \frac{\lambda_0-\lambda_1}{\lambda_0+\lambda_1}. \quad (13)$$

The weighted IPLD reflects the proportion of the primary component in the downmix signal. If the input signals contain no ambience, that is, if PTR is 1, then IPLD estimation is analogous to the conventional method. On the other hand, IPLD decreases as the power of the primary component decreases, which prevents the panning of ambient components.

3. Robust ICC Estimation Using ICTD Compensation

In the proposed algorithm, we adopt a robust ICC estimation which uses the interchannel time difference (ICTD) compensation previously proposed in [4], which aims to improve the conventional ICC estimation method using the constant interchannel phase difference (IPD) assumption.

The time/frequency mapping of the proposed algorithm is the same as that of MPS, which is configured as 20 parameter bands using a hybrid QMF filter. The ICC estimation algorithm in [4] applies to the frequency range from 3 kHz to 8 kHz, where parameter bands are composed of two or more hybrid sub-subbands.

4. Upmix Matrix to Reconstruct IPLD and ICC

The upmix matrix reconstructs two channel signals using the downmix and decorrelated signal. The proposed algorithm determines the elements of the upmix matrix so that they match the spatial parameters of the output signal to those of the input signals. Two different forms of upmix matrix are possible under two different assumptions, namely, anti-phase and orthogonal ambiences. Neither form made a perceptual difference in the informal listening test, while the latter doubled the computation load; therefore, we have adopted the former, which is formulated as

$$\begin{bmatrix} \hat{X}_1[k] \\ \hat{X}_2[k] \end{bmatrix} = \begin{bmatrix} l_1 \cos\alpha & (\sin\alpha)/\sqrt{2} \\ l_2 \cos\alpha & -(\sin\alpha)/\sqrt{2} \end{bmatrix} \begin{bmatrix} \tilde{S}[k] \\ D(\tilde{S}[k]) \end{bmatrix}, \tag{14}$$

$$\alpha = \arctan(\sqrt{2\beta}), \quad \beta = \frac{\gamma \pm \sqrt{\gamma^2 - 4l_1^2 l_2^2}}{2},$$

$$\gamma = \frac{2l_1 l_2 + ICC^2}{1 - ICC^2}, \; l_1 = \sqrt{\frac{10^{IPLD/10}}{1+10^{IPLD/10}}}, \; l_2 = \sqrt{\frac{1}{1+10^{IPLD/10}}}.$$

The process of driving this upmix matrix is similar to the conventional MPS [2].

IV. Performance Evaluations

1. Objective Simulation Results

A. IPLD Estimation

To parametrically evaluate the proposed coding algorithm, we first compare IPLD to CLD. The test signal was composed of male speech as the amplitude-panned primary component and ambient components of equal energy. We generated the ambient components by convolving the primary component with two exponentially decaying white Gaussian noises. Figure 1 shows the estimated parameters for the ninth and tenth parameter bands. The IPLD estimates follow the panning gain ratio more closely than the CLD estimates, especially when the panning direction is biased toward each channel's direction, $\pm 30^{\circ}$. Thus, we conclude that IPLD is a more appropriate parameter than CLD in terms of preserving the original panning gain ratio.

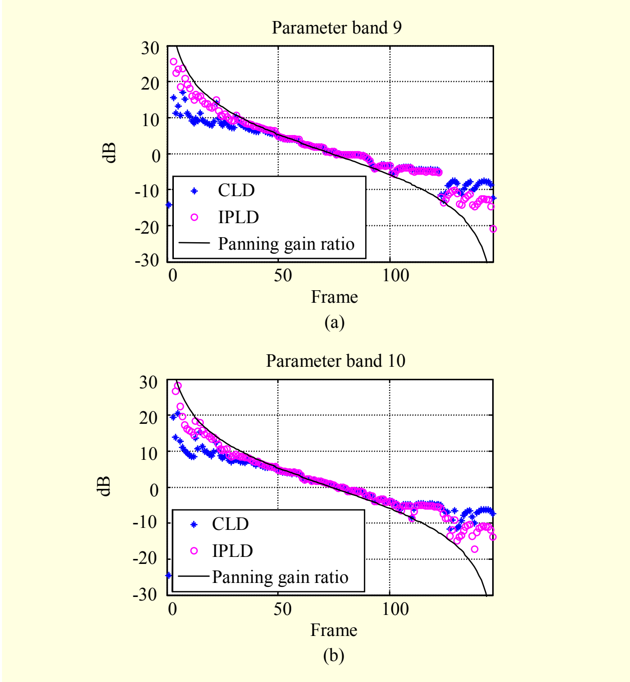

Fig. 1. Comparison of CLD and IPLD.

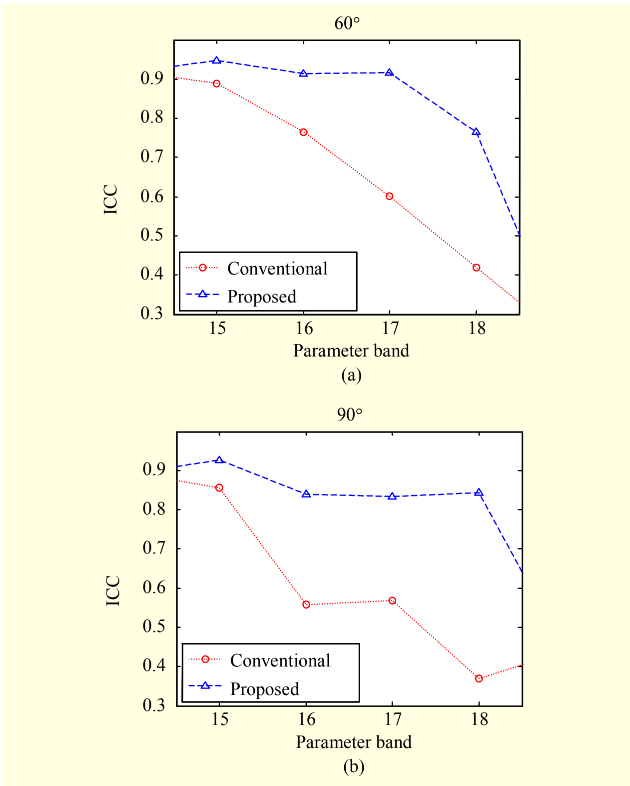

Fig. 2. ICC estimation results obtained by conventional (dotted) and proposed (dashed) methods.

B. ICC Estimation

To evaluate the robust ICC estimation in [4], we filtered white Gaussian noise with head-related impulse responses (HRIRs) corresponding to 60 and 90 degrees using the MIT

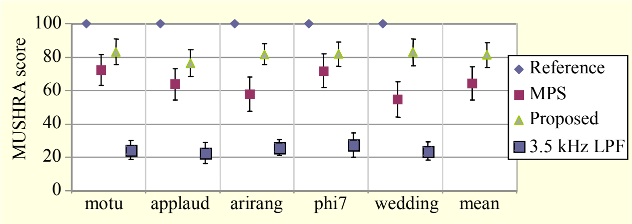

Fig. 3. Subjective listening test results.

HRTF Database [6]. Because the HRIR was recorded in an anechoic chamber, ICC should be 1 for all frequency bands. The averaged ICCs shown in Fig. 2 indicate that the conventional method tends to underestimate ICC. The new method, on the other hand, provides accurate ICC.

2. Subjective Listening Test

We conducted a MUSHRA test with seven experienced subjects under headphone playback [7]. The test compared conventional MPS to the proposed system using five stereo samples. The mean overall results and 95% confidence intervals of the rating are shown in Fig. 3. The results demonstrate that the proposed algorithm based on PCA always provided the same or better audio quality than MPS.

V. Conclusion

This letter proposes a new joint channel coding method based on PCA using active downmixing. The preservation of the PAR of the downmix signal is proven by mathematical equations. The limitations of PCA are overcome by using weighted IPLD instead of CLD. The results of computer simulations and subjective listening tests demonstrate that the proposed system offers better audio quality than MPS.

References

[1] J. Breebaart et al., "Parametric Coding of Stereo Audio," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 9, 2005, pp. 1305-1322.

[2] J. Breebaart et al., "Background, Concept, and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression," *J. Audio Eng. Soc.*, vol. 55, no. 5, 2007, pp. 331-351.

[3] C. Faller, "Parametric Multichannel Audio Coding: Synthesis of Coherence Cues," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 1, 2006, pp. 299-310.

[4] D. Hyun et al., "Robust Interchannel Correlation (ICC) Estimation Using Constant Interchannel Time Difference (ICTD) Compensation," *127th Conv. Audio Eng. Soc.*, 2009, (preprint 7934).

[5] M. Goodwin and J.M Jot, "Primary-Ambient Signal Decomposition and Vector-Based Localization for Spatial Audio Coding and Enhancement," *IEEE-ICASSP*, vol. 1, 2007, pp. 9-12.

[6] B. Gardner and K. Martin "HRTF Measurements of a KEMAR Dummy-Head Microphone," *Perceptual Computing Technical Report*, MIT Media Lab, no. 280, 1994. Available: http://sound.media.mit. edu/resources/KEMAR.html

[7] ITU-R BS.1534, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems (MUSHRA)," ISO, Geneva, Switzerland, 2001.